Transcriptional profiling of *Aspergillus niger*

Douwe van der Veen

Transcriptional profiling of *Aspergillus niger*

D. van der Veen

Transcriptional profiling of *Aspergillus niger*

Douwe van der Veen

Promotor

Prof. dr. J. van der Oost
Persoonlijk hoogleraar bij het laboratorium voor Microbiologie
Wageningen Universiteit

Copromotor

Dr. ir. L. H. de Graaff
Universitair docent bij het laboratorium voor Microbiologie
Wageningen Universiteit

Promotiecommissie

Prof. dr. H.A.B. Wösten
Universiteit Utrecht

Prof. dr. ir. F.P.M. Govers
Wageningen Universiteit

Prof. dr. J.T. Pronk
Technische Universiteit Delft

Dr. ir. M.J. van der Werf
TNO Kwaliteit van Leven, Zeist

Dit onderzoek is uitgevoerd binnen de onderzoeksschool VLAG

Transcriptional profiling of *Aspergillus niger*

Douwe van der Veen

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
Prof. dr. M.J. Kropff,
in het openbaar te verdedigen
op woensdag 17 juni 2009
des namiddags te half twee in de Aula.

Veen, Douwe van der
Transcriptional profiling of *Aspergillus niger*
Ph.D. thesis, Wageningen University and Research Centre, The Netherlands, 2009.
With summaries in English and Dutch.

ISBN: 978-90-8585-370-1

Table of contents

Abstract

The industrially important fungus *Aspergillus niger* feeds naturally on decomposing plant material, of which a significant proportion is lipid. Examination of the *A. niger* genome sequence suggested that all proteins required for metabolic conversion of lipids are present, including 63 predicted lipases. In contrast to polysaccharide-degrading enzyme networks, not much is known about the signaling and regulatory processes that control lipase expression and activity in fungi. This project was aimed to gain better understanding of lipid degradation mechanisms and how these processes are regulated in *A. niger*, primarily via assessment of its gene transcription levels.

Minimizing biological and technical variation is crucial for experiments in which transcription levels are determined, such as microarray and quantitative real-time PCR experiments. However, *A. niger* is difficult to cultivate in a reproducible way due to its filamentous growth. In addition, the complex processing steps of transcriptomics technologies add non-experimental variation to the biological variation. To reduce this data noise, robust protocols based on a batch-fermentation setup were developed. Variation in this setup was surveyed by examining the fungal transcriptional response towards a pulse of D-xylose. The sources of non-experimental variation were described by variance components analysis. Two-thirds of total variation stems from differences in routine handling of fermentations, but in absolute terms this variation is low. As D-xylose is an inducer of the xylanolytic system, the high reproducibility of cultures for the first time allowed a detailed description of the global fungal transcriptional response towards D-xylose using microarrays.

The transcriptional response towards three plant derived oils was examined in another study. Both olive oil and a wheat-gluten extracted oil induce the transcription of genes involved in lipid metabolism and peroxisome assembly, albeit with different expression profiles. The third oil, a plant membrane lipid, did not trigger a transcriptional response.

Microarray data are related to the physiology of the fungus. To better understand the general principles that underlie gene regulation and gene transcription, microarray data from cultures grown under mildly and strongly perturbed conditions were analyzed for co-expression of genes. Despite the diverse culturing conditions, co-expressed gene modules could be identified. Some of these modules can be related to biological functions. For some modules, conserved promoter elements were identified, which suggests that genes in these modules are regulated on a transcriptional level.

The work described in this thesis shows that (i) high-quality -omics data for *A. niger* can be generated; that (ii) analysis and interpretation of these data enhances our understanding of the xylanolytic and lipid metabolic regulons; and (iii) that these data give insight into the regulatory mechanisms of this fungus.

Chapter 1

Introduction

Douwe van der Veen

Introduction

On life

Biologists study life. While in most cases 'life' is readily recognized - humans are considered alive, but rocks are not - it has been surprisingly difficult to define this phenomenon. Throughout the ages, thinkers from diverse cultural and ideological backgrounds have tried to comprehend, explain, and define 'life'. However, life has many different facets at various levels of organization: consider for example a cell versus an ecosystem. To capture its essence in a single conceptual scheme has proven difficult.

The main purpose of a definition is to (i) demarcate or classify a certain phenomenon, or (ii) to make manifest - and perhaps even explain - the fundamental nature of that phenomenon [219]. Classification of 'life' leads mostly to descriptive lists of properties that discriminate living from non-living and may include such terms as complexity, natural selection, self-organization, or metabolism [72, 160]. The often-cited 'chemical Darwinian' definition: '*life is a self-sustained chemical system capable of undergoing Darwinian evolution*', falls into this category [1]. Two objections to using such classifying definitions can be made. The first objection is practical: how to make certain that no essential properties are left out or that non-essential properties are included? The second objection to classifying definitions is their inability to explain the underlying causes or the mechanisms leading to the production of these properties. That is, such definitions do not lead to a better understanding of the nature of 'life' [219].

The second approach to defining 'life' is to specify its meaning by building upon concepts that have been developed within a given theoretical framework. Cleland and Chyba [49] illustrate this approach by discussing definitions of 'water'. Water can be classified by reference to its sensible properties, such as it being wet, odorless, tasteless, or thirst quenching. However, this listing of properties allow substances that only superficially resemble water to be incorrectly classified as 'water'. The development of molecular theory made possible to describe 'water' unambiguously: water *is* H_2O. This conceptual definition within the theoretical framework of chemistry gives a broad, theoretically grounded, scientific understanding of 'water'. It is empirically testable, and allows to explain its properties and behavior under a range of chemical and physical circumstances, such as when in liquid or in solid phase.

[1] Note that this definition excludes a single specimen from being considered alive, as classic Darwinian evolution applies to a population only. For discussion, see [160].

While some philosophers argue that the absence of a theoretical framework of biology prevents defining 'life' [49], others argue that 'life' can be defined within the framework of particular theories already present (for overview, see [72, 160]). One elegant definition of 'life' is within the framework of cybernetics. Cybernetics is the study of the structure of regulatory systems, or '*the science of control and communication*' in both organisms and machines [13]. It deals with all forms of behavior insofar as it is regular, determinate, or reproducible – concepts all intuitively connected to life. A living individual can be regarded within the cybernetic theoretical framework as a system of regulatory mechanisms [145]. The aim of these mechanisms is to maintain the stable and constant identity of this individual within its physiological and biological environment. Deviations from this stability are corrected by negative feedback mechanisms. Such mechanisms are hierarchically organized and mutually interconnected, and exist at different levels. For example, at the biochemical level, an end product may block the pathway of its own production when present in high concentrations. At the genetic level, a gene's transcription rate is the outcome of concerted action of a variety of transcription activators and repressors. And at cellular level, genome-wide oscillations in transcription affect the physiological state of cells [143]. The purpose (in the biological sense) of this system of regulatory mechanisms is to be an effective replicator. That is, the system aims to copy and multiply itself, which is a positive feedback mechanism. Thus, using cybernetics concepts, 'life' can be defined as '*a system of inferior negative feedbacks being at service of a superior positive feedback*' [146].

Molecular biologists attempt to understand the complexity of life from the perspective of the macromolecules that life itself generates, such as proteins or DNA. For them, these are fascinating times. The field of molecular biology has been changed especially in the last decades by fast-paced technological developments. For example, rapid DNA sequencing technologies have revealed the complexity and variety of genes and genomes; advances in -omics technologies have allowed for taking snapshots of cellular components in a variety of situations; and bioinformatics has connected and facilitated analysis of molecular physiology data. These advances in experimental biology have enabled biologists to study genes, proteins, or biochemical reactions in great detail. In addition, they have made possible studies of their relations and dynamics [40, 140, 155]. Current attempts towards understanding life's complex biological organization are described in the often used but diversely defined catchphrase, 'systems biology'.

Two different interpretations of 'system' underlie the use of 'systems biology' [192] - and these interpretations are not mutually exclusive. The first interpretation is given by researchers who find it useful to refer to the particular phenomenon that they study as a 'system'. For these pragmatic systems biologists, 'systems biology' is a convenient notion that refers to classical biological experimentation but on a larger scale. The use of

new technologies generates a wealth of data, which enables them to extend their analyses to the entire phenomenon they study - in contrast to the technology-limited inquiries of the past [1, 14, 38]. A second interpretation of 'systems biology' comes from systems-theoreticians, who insist that this pragmatic approach to investigating a system is inadequate. For these biologists, it is crucial to analyze systems as systems and not as mere collections of parts only. Interactions between individual parts may result in distinct, collective, and interactive properties that do not and cannot manifest themselves in these individual parts alone. Such emergent properties must be described as well [31, 51]. In this interpretation, a system is viewed as a fundamental category where the similarities between biological and engineered systems are more important than their differences.

A paradox in current molecular biology is that, while technological advances allow to gather vast amounts of data on an increasingly number of biomolecules, little in the way of general theory has emerged that is able to explain how emergent properties arise from the biomolecules that are present in the studied system. The absence of such theoretical framework is not necessarily a problem: biologists are adept at exploring 'life' and conducting 'systems biology' without a pressing need to define these concepts. This is exemplified by the wide body of knowledge available: the PubMed search engine covers over 18 million citations from biomedical literature [186], and in the field of molecular biology currently 1,170 scientific databases are available to researchers [82]. Nevertheless, a conceptual clarification can be of practical importance for biologists. Carl Woese remarked on the development of science, that "*a properly balanced relationship between technological advances and a guiding vision is key to the successful development of a science: without the proper technological advances the road ahead is blocked. Without a guiding vision, there is no road ahead.*" [288]. Definitions of concepts enable the direction of a scientific experimental program or propose the orientations through which biological data should be approached [2].

[2] The relation between a definition of 'life' and experimental work is intuitively more clear in more exotic areas of biology. In astrobiology, implicit or explicit definitions of 'life' guide life detection strategies [50]. The Viking mission to planet Mars included equipment to detect metabolizing organisms - thus assuming that one characteristic of 'life' is the capacity to metabolize, *i.e.* incorporate carbon dioxide or carbon monoxide into organic compounds [141]. Research on prebiotic life, pushed to modern science by Alexander Oparin's thesis that the transition to life is a process which goes spontaneously from the inanimate matter to the first living cells [194], also requires a definition of 'life' prior to experimentation.

Between simple and complex, a thin line

Implicit in the work described in this thesis is the notion that life's complexity eventually can be described by models that are able to capture the dynamic behavior of interacting biological parts. Manifestations of life are often cited as supreme examples of complexity. It is a common assumption that this observed complexity is the result of underlying principles that themselves are complex. However, this assumption is not necessarily correct. A complex situation can arise from simple rules, while a seemingly simple situation can be orchestrated by complex underlying processes. Two examples serve to illustrate this in the sections below.

That simple rules can lead to complex outcomes is demonstrated by the behavior of cellular automata [20, 290]. Cellular automata gained popularity in the 1970s after

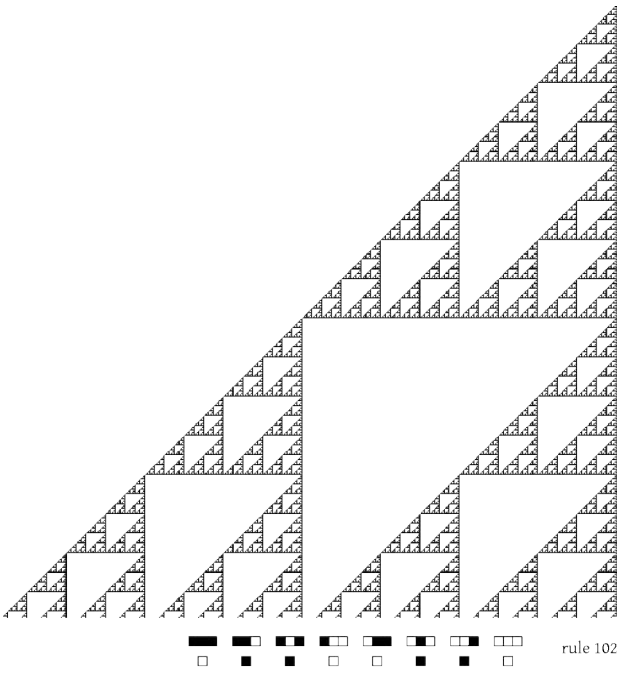
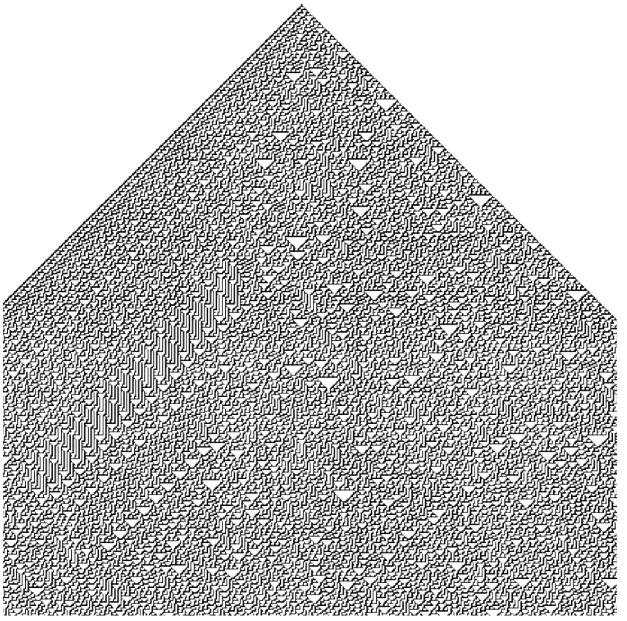
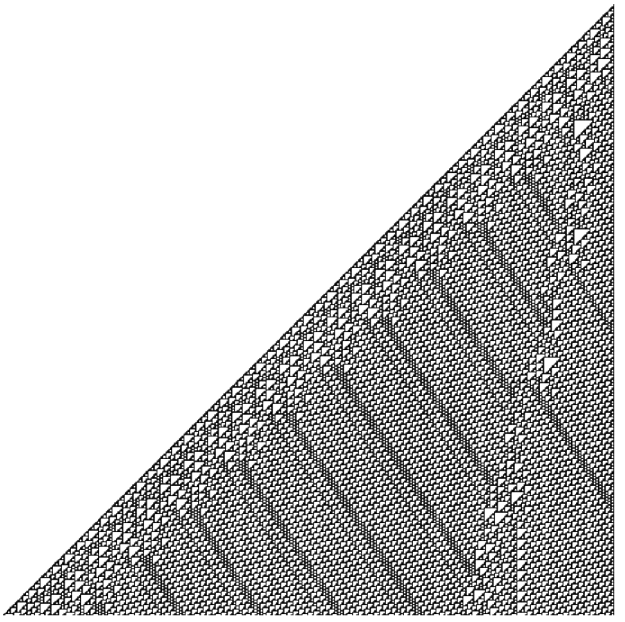


Figure 1 (continued on next page) - Three rule sets describing the color of cells of three one-dimensional cellular automata, and a visual representation of 600 time steps for each automaton. Under each automaton, the legend describes the eight possible combinations of a cell and its neighbors on the top, and the resulting cellular state of that cell for the next time step on the bottom. For example, for rule 102, when a black cell has to its left a white cell and to its right a black cell, its color turns to white in the next time step. Rule names are derived from their binary representation following [290]: *e.g.* decimal value 102 has binary value 01100110.



rule 30



rule 110

publication of a series of articles on John Conway's '*Game of Life*' [3] [83], but are used professionally to describe such processes as cell movement [100] or immune system responses [174]. A cellular automaton is a mathematical model that is easily visualized as a regular grid consisting of squares, or cells (not to be confused with biological cells). Cells are assigned a certain value; for the simple cellular automata discussed here, a cell can be either black or white (Figure 1). The state of every cell is updated in discrete time steps according to predefined rules. One such rule can be that "a black square with both its left and right neighbors being black, becomes white at the next time point". Three one-dimensional cellular automata, each based on a different set of rules, are depicted in Figure 1.

While the rule sets of these automata are similar, their outcomes are not. It is obvious that a regular pattern emerges when the so-called rule 102 is applied. Triangle-shaped pieces of identical form appear and it can be seen in Figure 1 that each piece is essentially just a smaller copy of the whole pattern. The opposite of order is apparent when rule 30 is applied. This rule set produces an irregular and complex pattern, and it can mathematically be shown that this rule creates perfect randomness [90, 290]. However, in essence this perfect randomness also has its order: randomness dictates that on average the colors black and white occur equally often. Execution of a third rule, rule 110, produces a pattern that is neither completely random not completely repetitive but is a mixture of both. Interestingly, this rule only differs from the regular pattern-obtaining rule 102 by one digit (Figure 1).

The above-mentioned visualizations illustrate that simple rules can generate complex patterns that range from full randomness to full regularity. The regulatory mechanism of the *Saccharomyces cerevisiae* galactose utilization (GAL) pathway, one of the most intensively studied and well understood genetic regulatory circuits, serves as example for the reverse situation where apparent simple behavior is obtained by interacting sets of complex interactions.

S. cerevisiae can utilize galactose using the enzymes of the Leloir pathway. These enzymes convert galactose to glucose-6-phosphate which in turn enters the glycolysis pathway (Figure 2) [78]. In the canonical model, expression of the genes encoding these enzymes is tightly regulated by pathway-specific regulators as well as by global carbon catabolite repression [127, 157]. Growth of yeast on glucose completely represses the transcription of the GAL genes due to carbon catabolite repression. But in the absence of glucose and presence of galactose, three regulatory proteins act co-operatively to

[3] An online version of the "Game of Life" can be found at http://www.bitstorm.org/gameoflife/

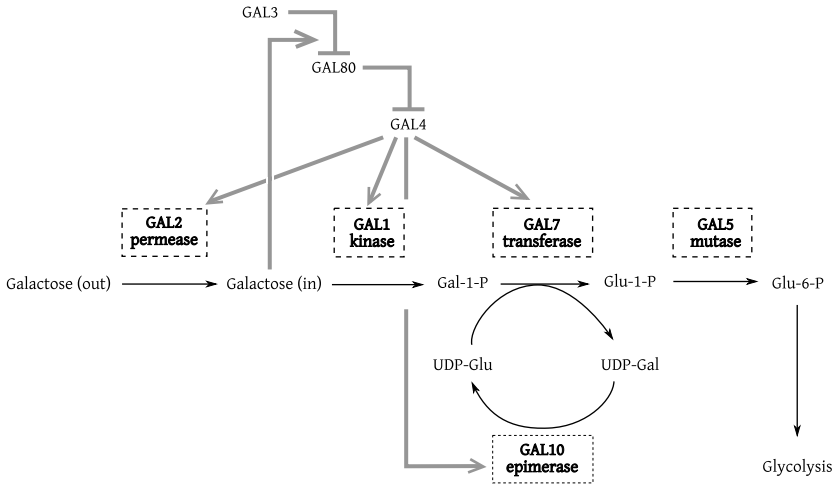


Figure 2 - The galactose utilization (GAL) pathway and its core regulators. Galactose is transported into the cell by the permease encoded by GAL2 as well as other hexose sugar transporters (not shown). The enzymes are galactokinase (encoded by GAL1), galactose-1-phosphate uridylyltransferase (GAL7), uridine diphosphoglucose 4-epimerase (GAL10), and phosphoglucomutase (GAL5). Grey arrows indicate regulatory mechanisms. Intracellular galactose is sensed by GAL3, which binds to GAL80 and thus abolishes its inhibition of GAL4. GAL4 is the transcription activator for GAL1, GAL2, GAL7, and GAL10. Although GAL5 expression is weakly induced by GAL4 upon galactose induction, this gene is constitutively expressed [80]. Figure based upon data in [122, 127].

express the genes of the GAL pathway. The GAL4 protein is a DNA-binding transcription factor that binds to upstream sequences of the functional enzyme-encoding genes GAL1, GAL2, GAL7, and GAL10, and thus activates their transcription. When no galactose is present inside the cell, the GAL80 protein interacts with GAL4 and inhibits its activity. GAL3 is the ligand sensor of the regulatory circuit. Upon sensing the presence of both galactose and ATP in the cell, GAL3 interacts with GAL80 [201, 254]. This association causes GAL80 to release its repression of GAL4, so that the genes encoding the functional enzymes of the Leloir pathway are expressed at a high level.

It was known that the complete regulatory mechanism was more complex than the core regulation described above [197, 214]. However, a systematic series of perturbations of the GAL system that were examined by large-scale mRNA expression and protein measurements revealed a complex picture of this regulon [122]. In this study, wild-type yeast and nine GAL mutants were grown in the presence or absence of galactose, and their transcriptome and proteome were analyzed. Some observations were made that were not predicted by the model of Figure 2. For example, in the presence of galactose,

deletions of either the GAL7 or GAL10 genes reduced the expression levels of other GAL enzymes. This observation suggests a regulatory role for either the metabolite galactose-1-phosphate or one of its derivatives, or a regulatory function of the GAL10 enzyme [122]. This systematic approach also revealed that mRNA levels of 997 genes differed significantly in at least one such perturbation - much more than the limited number of genes thought to be affected prior to this study. Integration of these data and its coupling with protein-protein and protein-DNA interactions data showed that processes not directly related to galactose metabolism, such as amino acid synthesis or fatty acid oxidation, are also influenced by perturbation of the GAL system.

Fungal life

Up to this point, this introduction has touched briefly on concepts and challenges towards investigation life's complexity in general terms. Focus now shifts towards the study of fungal life with emphasis on *Aspergillus niger*, the fungus on which the work described in this thesis is based.

Fungi are eukaryotic micro-organisms. An estimated 1.5 million fungal species [101] inhabit a wide range of ecosystems on earth, where these organisms play a vital role [105]. White rot fungi and brown rot fungi are the most efficient degraders of lignocellulose, a major component of wood or straw [32]. The majority of terrestrial plants live in association with symbiotic fungi that facilitate the uptake of mineral nutrients [223]. Edible fruiting bodies of fungi such as shiitake, champignon, or truffle serve as food for humans (while ants prefer fungi of the order *Agaricales* [53]). Food processes including wine making, beer brewing and bread baking involve fungal activity. Penicillin, blood pressure lowering agents, and other pharmaceuticals are of fungal origin [108]. As lower eukaryotes, their cells function in many ways similar to a human cell, which makes some species important model organisms in medical research. On the negative side, some fungi display undesired properties: the major plant pathogens are fungi and several hundred fungal species cause disease in animals or man [22].

Aspergillus niger is a black sporulating mold whose physiological properties have made it a workhorse of modern biotechnology. Its natural habitat is soil, an environment where nutrients are often scarce and competition is fierce. Here, it feeds primarily on dead plant matter. As a majority of bacteria and many fungi cannot grow below pH 3, by acidification of its environment *A. niger* inhibits the growth of competitors and increases its own access to nutrients [167]. To this end, it can produce large quantities of organic acids. This characteristic is utilized in industry, where this fungus has become the almost exclusive industrial producer of citric acid worldwide.

A. niger is also an excellent producer of extracellular enzymes. These enzymes are aimed primarily to degrade plant matter - again, a capacity that is tailored to its natural environment. A multitude of specialized enzymes are secreted into its surroundings where their activities can convert complex biomolecules into consumable substrates. Prime examples of such enzymes include pectinases and xylanases. Even though the primary function of these enzymes are to degrade plant matter, they are used in baking, brewing, and beverage industries, in animal feed and in the paper and pulp industry [196]. In addition to its own enzymes, *A. niger*'s excellent protein secretion machinery is exploited to produce a range of heterologous proteins [207].

Post-genome research of *Aspergillus niger*

Before an integrated approach to understanding the physiology of *Aspergillus niger* can be undertaken, the individual biological parts contributing to its complexity need to be identified. The complete genome sequences of two *A. niger* strains are publicly available and provide an excellent starting point for this inventory. A genome harbors the chromosomal segments that are responsible for making a functional product, either a protein or non-coding RNA [68, 240]. The genome of strain CBS 513.88, a predecessor of currently used enzyme production strains, is 33.9 Mbp in size and contains 14,165 protein-coding genes [196]. Citric acid producing strain ATCC 1015 has a slightly larger genome size of 34.9 Mbp but contains 11,197 protein-coding genes [128]. However, information from ongoing or completed genome sequencing projects suggest that genomes by no means are ordered and logical libraries of biochemical instructions for making a cell. Instead, they are the imprint of occurrences in evolutionary history, selection, and biochemical constraints: hence pseudo-genes, paralogs, transposons, or DNA sequences with no apparent functional benefit.

In addition, an organism's emergent properties do not arise from the information about the synthesis of its numerous biomolecules alone but from their combined activity and interactions. Variations in the timing, location, and levels of gene transcription, protein synthesis or metabolism have considerable consequences for a cell. It therefore is not sufficient to only investigate *A. niger*'s genome when trying to understand its physiology. Technologies built upon and using genome sequence data have been developed that enable the study of these spatio-temporal dynamics for a number of biomolecule classes. Transcriptional events can be assessed by transcriptomics technologies such as DNA microarrays [26, 36]. The protein content of organelles or whole organisms, protein-protein interactions, or the spectrum of secreted proteins can be mapped using proteomics technologies [137, 138]. And the continuum of metabolites - low molecular weight organic compounds with molecular weights of less than 1000 Da - can be surveyed using metabolomics [265].

The quantification of gene expression at the mRNA level is one of the most extended techniques in molecular biology research. The first microarray developed monitored the expression of only 45 *Arabidopsis thaliana* genes [225], but two years later, all coding genes of the yeast *S. cerevisiae* were dotted onto a 1.8-cm-square glass slide to analyze its global gene expression levels [150, 231]. Currently, microarrays are routinely and widely used within almost all areas of biological research. Whole-genome microarrays have been developed for seven species of filamentous fungi, including for *A. niger* [4]. For this thesis, the availability of an *A. niger* microarray platform provided an excellent starting point to study the transcriptional response of this fungus in a variety of experimental settings. However, the development and use of -omics technologies brought to the spotlight two closely related experimental challenges.

The first challenge relates to the technical side of -omics technologies. Gene transcript levels currently can be determined with great precision and accuracy: as little as 20 absolute mRNA copies can be detected by microarrays [233]. This sensitivity, in combination with its ability to simultaneously determine the expression profiles for many genes, demands that non-experimentally induced variation ('random noise') be kept to a minimum. The effects of this random noise on data analysis and data interpretation can be minimized by the use of replicate biological samples to generate high quality data. However, the second experimental challenge lies in this biological reproducibility. The filamentous or 'thread-like' growth of *A. niger* deters the growth of reproducible cell cultures. The fungal filaments increase the viscosity of the culture medium during cultivation, which in turn leads to a reduced oxygen and heat transfer ultimately resulting in a heterogeneous cell culture. Physical agitation and shear stress contribute to this heterogeneity by causing uncontrolled breakage and fragmentation of the mycelium.

Outline of this thesis

The work presented in this thesis is the result of a joint partnership of DSM and Wageningen University. The initial project objective was to investigate lipid degradation and metabolism in *A. niger*. However, over time, the experimental requirements inevitably lead to a broader focus. This work can be viewed from two connected perspectives, each of which emphasizes a different aspect on the study of the complexity of *Aspergillus niger*.

[4] For four filamentous fungi, two arrays per species were developed independently: for *A. niger* (DSM N.V., The Netherlands, and [11]), *A. nidulans* ([11, 29]), *A. fumigatus* (J. Craig Venter Institute, USA, and [188]), and *A. oryzae* (JCVI and [11]). One whole-genome array was developed for *A. flavus* (JCVI), *Fusarium graminearum* ([92]), and *Neurospora crassa* (Agilent Inc., USA).

The first perspective on the work described here is on the acquisition of high-quality experimental data. As mentioned above, the culturing and processing of biological material introduces non-experimental variation. The resulting data noise affects the quality and reduces the reliability of the data obtained. This works describes efforts to minimize and control this non-experimental variation. In **Chapter 2**, microarray data are used to discover genes that have stable expression levels under a wide variety of conditions. Such endogenous reference genes, or 'housekeeping genes', are required for normalization of quantitative real-time PCR data, a method to analyze transcript levels of a single gene. The resulting method to accurate determine transcript levels was used in **Chapter 3** to assess the improvements in the reproducibility of fermentor-grown *A. niger* cell cultures. The procedure outlined in this chapter is a fast and effective way to determine the source and magnitude of non-experimental variation introduced during growth experiments.

The second perspective on this work views the contributions towards understanding the biology of *A. niger*. The ability to grow highly reproducible cell cultures of *A. niger* and the ability to acquire high-quality gene expression data from these cultures allowed for a detailed analyses of transcriptome fluctuation. In **Chapter 3**, the fungal global transcriptional response towards D-xylose is examined. This sugar is an inducer of the xylanolytic system, and the transcriptional response towards it was known to some extent. However, for the first time, the global analysis of its response towards D-xylose could be described in great detail.

In **Chapter 4**, the transcriptional response towards a pulse of three oils – olive oil, a complex oil mixture extracted from wheat gluten, and the plant membrane lipid DGDG - is described. *A. niger* feeds naturally on decomposing plant material, a significant proportion of which is lipid. In contrast to plant cell wall degrading enzyme networks, not much is known about lipid degradation of *A. niger*. This study showed that both olive oil and the complex oil mixture induce the transcription of genes involved in lipid metabolism and peroxisome assembly, and they do so with different expression profiles. No transcriptional response towards DGDG was observed.

Microarray technology data are related to the physiology of the organism. In Chapter 3 and 4, there was an emphasis on data analysis to understand a specific reaction towards inducer compounds in greater detail. The data obtained in these studies are combined with microarray data of *A. niger* cultures grown under strongly perturbed culture conditions in **Chapter 5**. This chapter underscores the strength of re-analysis of complex data sets. The aim of this analysis was to gain insight into general principles that underlie transcript expression and gene regulation. A gene co-expression analysis was used that resulted in the identification of modules of co-expressed genes, despite

the use of distinct culturing conditions. Examination of genes present in these modules indicates an overrepresentation of biological functions per module. Also, conserved promoter elements were identified in the upstream region of its genes, suggesting co-regulation of genes in these modules on a transcriptional level.

The availability of more and more precise data such as generated in this thesis greatly enhances the understanding of biological parts and their interactions. However, for a large proportion of proteins, a physiological function is still unknown. A method proposed in **Chapter 6** aims to make use of expression data to infer gene function by following the 'guilt by association' principle. This principle assumes that unknown genes that are co-expressed with genes of known function most likely encode proteins that are involved in the same biological process. Although this method does not reveal the precise biochemical function of the unknown protein, it provides valuable starting points for further investigation.

Up to date, no theoretical framework or models have been constructed that integrate multi-level biological and biochemical data to describe the full functioning of a living organism. Certainly, it is worth striving for, and facilitating the development of, such integrative models. Life is a complex phenomenon, and to study the dynamics and interactions of its biomolecules is challenging. This thesis demonstrates that high-quality data for *A. niger* can be generated, and it demonstrates that analysis and interpretation of these data enhances our understanding of the biology of *A. niger*. In addition, it suggests that analysis of the regulatory mechanisms of these processes aids in unraveling the complexity of this fungus. Moreover, set apart from work described here, it acknowledges that biologists are just barely starting to understand the phenomenon of 'life'.

Chapter 2

Selection of *Aspergillus niger* endogenous reference genes using microarray data, and comparison with a synthetic RNA reference

Douwe van der Veen, José Miguel Oliveira,
Elena S. Martens-Uzunova, and Leo H. de Graaff

Abstract

Real-time quantitative PCR (qPCR) has become a standard technology for determining transcription levels. Before analysis of the data obtained by qPCR, these data have to be corrected for experimental variation. This correction, or data normalization, is commonly done using internal reference genes – genes believed to be stably expressed under the physiological conditions of interest. However, this strategy suffers from drawbacks as the transcription levels of commonly used reference genes such as β-actin or GAPDH do vary under many conditions. Therefore, the stability of reference genes selected for use in an experiment must be validated before their data can be used to normalize qPCR data. In this study, microarray data were analyzed to identify stably expressed genes beyond the realm of traditionally established reference genes. An alternative for internal reference genes, spiking by use of a synthetic RNA transcript, was investigated as well.

Data of 48 *Aspergillus niger* microarrays were analyzed and 321 candidate reference genes were identified. The transcript expression stability in qPCR measurements was assessed for 11 candidate genes in a novel growth experiment using a different *A. niger* strain and a distinct substrate as a sole carbon source. The candidate reference genes were less variable with respect to their transcript levels as compared to traditional reference genes proposed for *A. niger.* A synthetic RNA transcript was spiked into the total RNA pool and evaluated for use as exogenous reference gene. This synthetic gene outperforms the tested candidate reference genes in terms of minimal variability between samples.

Large-scale expression data mining is a valuable tool to identify more stably expressed candidate reference genes compared to traditionally used reference genes. The use of a synthetic RNA spike enhances the reliability of qPCR normalization, and serves as detector for aberrant behavior of used endogenous reference genes in the qPCR samples.

Background

Although the principles of the polymerase chain reaction (PCR) were described as early as 1971 [142], improvement of the PCR reaction by Kary Mullis and co-workers by the use of a thermostable *Thermophilus aquaticus* DNA polymerase in the 1980s has made it an indispensable technique for current molecular biology laboratory work [182, 220]. The basics of a PCR reaction are that two sequence-specific DNA oligonucleotides are used to amplify a specific DNA fragment. Addition of intercalating fluorophores to the reaction mixture allow the real-time monitoring of this amplification, and enable the quantification of the amount of DNA formed [106]. This real-time quantitative PCR technique was extended further by Gibson and co-workers, who used cDNA as template and thus made possible the measurement of gene expression levels [86]. Currently, real-time reverse-transcribed quantitative PCR (qPCR) has become a routine technology for the accurate quantitation of gene expression and is widely used for validation of DNA microarray results [205].

The reliability of this technology depends on data normalization to obtain comparable results. Such normalization is necessary to compensate both for biological variation and for differences in technical processing and treatment of these samples [41, 274]. Common practice in qPCR normalization is the use of internal control genes, also called housekeeping genes or endogenous reference genes [246]. A prerequisite for any gene to be considered an endogenous reference gene is that its transcription levels do not vary in the experimental conditions under study. Discussions on which genes to use as endogenous reference gene date from the Northern blot era [18] and continue until this day [33, 41, 104]. Unfortunately, reference genes that are selected in one condition for stable expression may vary in a next [62]. Current best practice is to validate a set of endogenous reference genes for stable expression, for instance by use of algorithms implemented in the programs geNorm [274], BestKeeper [199], or Normfinder [9].

Alternative normalization strategies have also been suggested. Normalization can be based on the starting amount of cells or the amount of total RNA used. However, this does not take into account the varying efficiencies of the reverse transcription or the PCR reaction itself [114]. Spiking of total RNA with alien RNA, as is performed routinely for microarray techniques, has been validated for qPCR [87] but has not yet been widely adopted. Genome-wide expression data for many organisms and experimental conditions is available in repositories such as GEO [69] and ArrayExpress [34]. Expression levels for thousands of genes are available, and this data can aid in the selection of endogenous reference genes for the organism of interest. So far, few reports investigate the feasibility of this approach [152].

The aim of our present study was to identify genes of the fungus *Aspergillus niger* that are stably expressed under various conditions, and to investigate their use in qPCR data normalization. Furthermore, we enhanced the robustness of our qPCR method by combining normalization strategies. In particular, we sought to minimize the erroneous effects of inappropriately chosen reference genes through the use of a synthetic exogenous reference transcript [61].

Results

Selection of candidate reference genes

A total of 48 *A. niger* microarrays was available in our laboratory at the onset of this study (Table 1). These data were obtained from several independent experimental setups and differed by (i) growth condition, (ii) media composition, (iii) sampling time point, and (iv) experimenter. These arrays were split into two groups. A selection group of 21 arrays contained samples obtained under mild conditions. We assumed that these samples would not show global stress responses due to the limited time elapsed after transfer, the excess of nutrients in the fresh medium, and the buffer capacity present in this medium. The remainder of 27 arrays were used to validate the candidate reference genes. Since these arrays originate from purposely imposed stressful conditions such as nutrient depletion or prolonged cultivation, large-scale fluctuations in transcript profiles are expected for this group.

Reliable probe sets were selected on the basis of their "present" flags and raw signal values as described in the Methods section. Of the 14,554 probe sets present on the *A. niger* array, 2,429 probe sets fit these criteria. We selected genes with invariable expression using the standard deviation [152]. Probe sets were excluded when the standard deviation of their expression values was above the median standard deviation for the Affymetrix control probes (mean SD = 0.20). This selection left 541 genes. Genes expressed over the fold change threshold for technical replicates (1.7-fold) were excluded, and a list of 469 candidate endogenous reference genes was selected.

The 469 candidate genes were selected from samples cultured under mild conditions. To investigate transcript level stability at non-favorable conditions, transcript levels were

Table 1 (opposite page) - Characteristics of DNA microarrays used in this study. Method: the cultivation method, which is either (S)hakeflask, (F)ermentor, or (P)aper. Time: for paper and shakeflask experiments, the time after transfer is given. For fermentor experiments, the time after start of fermentation is given.

Strain	Carbon source	Method	Time	Group
N400	D-galacturonic acid	S	2	selection
N400	D-galacturonic acid	S	4	selection
N400	D-galacturonic acid	S	8	selection
N400	D-xylose	S	2	selection
N400	D-xylose	S	4	selection
N400	D-xylose	S	8	selection
N400	fructose	S	2	selection
N400	fructose	S	4	selection
N400	fructose	S	8	selection
N400	D-galacturonic acid	S	2	selection
N400	D-galacturonic acid	S	4	selection
N400	D-galacturonic acid	S	8	selection
N400	rhamnose	S	2	selection
N400	rhamnose	S	4	selection
N400	rhamnose	S	8	selection
N400	sorbitol	S	2	selection
N400	sorbitol	S	4	selection
N400	sorbitol	S	8	selection
N400	sugarbeet pectin	S	2	selection
N400	sugarbeet pectin	S	4	selection
N400	sugarbeet pectin	S	8	selection
NW219	fructose	S	4	this study
NW219	fructose	S	4	this study
NW219	fructose	S	6	this study
NW219	fructose	S	6	this study
NW219	fructose	S	8	this study
NW219	fructose	S	8	this study
NW219	olive oil	S	4	this study
NW219	olive oil	S	4	this study
NW219	olive oil	S	6	this study
NW219	olive oil	S	6	this study
NW219	olive oil	S	8	this study
NW219	olive oil	S	8	this study

Strain	Carbon source	Method	Time	Group
NW219	wheat diglyceride	S	4	this study
NW219	wheat diglyceride	S	4	this study
NW219	wheat diglyceride	S	6	this study
NW219	wheat diglyceride	S	6	this study
NW219	wheat diglyceride	S	8	this study
NW219	wheat diglyceride	S	8	this study
N400	D-galacturonic acid	S	24	validation
N400	D-xylose	S	24	validation
N400	fructose	S	0	validation
N400	fructose	S	0	validation
N400	fructose	F	16	validation
N400	fructose	F	16	validation
N400	fructose	F	20	validation
N400	fructose	F	20	validation
N400	fructose	S	24	validation
N400	paper	P	6	validation
N400	paper	P	6	validation
N400	paper	P	18	validation
N400	paper	P	18	validation
N400	PD-galacturonic acid	S	24	validation
N400	rhamnose	S	24	validation
N400	sorbitol	F	20	validation
N400	sorbitol	S	24	validation
N400	sorbitol + polygalacturonic acid	F	16	validation
N400	sorbitol + polygalacturonic acid	F	20	validation
N400	sorbitol + polygalacturonic acid	F	20	validation
N400	sorbitol + sugarbeet pectin	F	16	validation
N400	sorbitol + sugarbeet pectin	F	20	validation
N400	sorbitol + sugarbeet pectin	F	20	validation
N400	sorbitol + sugarbeet pectin	F	20	validation
N400	sorbitol + sugarbeet pectin	F	40	validation
N400	sorbitol + sugarbeet pectin	F	40	validation
N400	sugarbeet pectin	S	24	validation

investigated on 27 microarrays of stress-imposed samples. Under these conditions, indeed transcript variability is higher. Four of 469 candidate genes are over 30-fold differentially expressed on some of the arrays. Unexpectedly, the 8 Affymetrix control probe sets included in the list of 469 genes follow these 4 genes in the next highest standard deviations on this set of arrays. A 2.3-fold cut-off value as criterion for transcript levels variability was used instead. This criterion was chosen on the basis that, in stringently controlled batch fermentations, genes do not differ by over 2.3-fold between biological replicates. Application of this cut-off value removes one-third, or 140, of the candidate genes and left 321 genes.

Functional classification of candidate reference genes

A combination of tradition and literature research commonly guides the composition of a pool of reference genes that are tested for suitability. Knowledge of the functional role of a protein encoded by a reference gene can be helpful in this process. To further refine our list of candidate reference genes, we used an in-house built database (Cluster of Orthologous Genes, or COG, database) [S. Basmagi and P. Schaap, unpublished data]. This database is based on publicly available yeast and fungal protein sequences. A protein clusters in a COG when corresponding ortholog sequences are detected in at least three organisms. The presence of a protein in such cluster is indicative of evolutionary conservation between species and increases the likelihood that the protein is functional. Two-thirds of the proteins that are present in a COG are highly conserved in the fungal kingdom as these have orthologs in over 15 organisms. In all but a few cases a *Saccharomyces cerevisiae* ortholog is included in these conserved clusters. Fifty-six proteins cluster in less conserved COGs and appear to be present only in ascomycetes. Genes were excluded from the candidate list when encoding proteins have no ortholog in the COG database (36 excluded), or for which no annotation information is available (32 excluded), leaving 253 genes.

S. cerevisiae has been used for many decades as a model organism and a wealth of information for it is available. Three quarters of the candidate gene list (187 genes) have a *S. cerevisiae* ortholog. The stringent criteria on which the COG database is built allow for transfer of information between proteins in the same COG cluster. Functional annotation of the 187 genes was performed by deriving information about their *S. cerevisiae* orthologs from the Gene Ontology Slim Mapper (*Saccharomyces* Genome Database [67]) and subsequent transfer of the annotations to their *A. niger* counterparts. The top five GO Slim processes show involvement in organelle organization and biogenesis (36%), transport (30%), vesicle-mediated transport (20%), protein modification processes (19%), and membrane organization (12%). The majority of proteins encoded by the candidate genes are predicted to be localized in the cytoplasm (68%), followed by localization in nucleus (34%), membrane (25%), mitochondrion (14%),

or the endoplasmatic reticulum and Golgi apparatus with 11% each.

Based on evolutionary conservation according to the COG database, *S. cerevisiae* ortholog information on cellular function and localization, and the raw expression signal values and their deviation across conditions examined by microarrays, a set of 11 candidate genes was selected for further analysis (Table 2).

Gene	Description	Signal	Primers
An11g11300	*A. niger* HttA, histone H2A	6300	F 5'-TGAACAAGCTCCTGGGTCAT R 5'-CTTTCCACTCTTGGGGGTCT
An08g06760	similarity to *S. cerevisiae* Sfi1 spindle pole body protein	110	F 5'-TGGACTGTATGCTCGGTCTG R 5'-ACTGATGCTGTCGTCACTGG
An08g07470	similarity to *R. norvegicus* VAMP-7 vesicle-associated membrane protein-7	280	F 5'-TTAGGAGTCGTGGCTTGAGG R 5'-TTGGGATACCCCAGGAGTAAA
An14g05050	strong similarity to *S. pombe* Dma1 defective in mitotic arrest	610	F 5'-ACTCCAGAGGACAAGCAGGA R 5'-GCAGACGCATGCTCTCAATA
An08g06940	strong similarity to *A. nidulans* histone H4.1	5100	F 5'-ATCTTGCGTGACAACATCCA R 5'-CACCCTCAAGGAAGGTCTTG
An12g07380	strong similarity to *A. parasiticus* MoxY monooxygenase	140	F 5'-AAACAGATATCCCGGCTGTG R 5'-CAGCTTCCATAACTTGCTGCT
An18g01970	similarity to *C. glutamicum* PanC pantoate-beta-alanine ligase	170	F 5'-CAAGCGTTTGGTGAAGGACT R 5'-GCCTTCATGGCTTCGTAGAG
An14g03180	strong similarity to *S. pombe* Cdc3, profilin	2600	F 5'-GGTATCCTGAGCCACGACAT R 5'-GCCTTGATGGTGACGAACTT
An09g06840	strong similarity to *S. cerevisiae* Sly41 involved in ER to Golgi transport	780	F 5'-ACCCCTGTTCACAGTTCTGG R 5'-CGCACATAGTATGCCGAAGA
An08g05910	strong similarity to *A. nidulans* SagA involved in cell transformation	150	F 5'-CCAGGATGAAGAGTGGGAGA R 5'-GCAGCTGGAGTGCTTCTTTC
An02g04120	similarity to *S. cerevisiae* Atx2 ZIP zinc transporter	500	F 5'-TTTTCAGTCTGGCTGCTCCT R 5'-CTGTTTTCCTGCATCGTGTG

Table 2 - The 11 candidate reference genes selected for qPCR testing. Gene, the *A. niger* gene code; Signal, average fluorescence intensity signal on Affymetrix microarray, in Arbitrary Units; Primers, the (F)orward and (R)everse primers used for qPCR.

Testing paradigm-dependence

The use of endogenous reference genes is paradigm-dependent, that is, transcript levels for an endogenous reference gene can be stable in one experimental condition but fluctuating in another [252]. Prior to the use of such genes as the basis for data normalization in a novel experimental condition, the stability of expression levels of these genes has to be validated. A new transfer experiment was used to verify the stability of the selected candidate reference genes under novel conditions. Wild type *A. niger* strain N400 was replaced by the low-conidiophore auxotrophic strain NW219. In

initial transfer experiments, candidate genes were selected based on on carbohydrate-grown samples only. Thus, for the new transfer experiment, two non-carbohydrate substrates were additionally chosen as sole carbon source: olive oil and an oil mixture containing mainly diglycerides. The transcript levels of the 11 candidate reference genes were determined by qPCR for fungal samples grown up to 12 hours after transfer to the new condition. The cycle threshold and amplification values were extracted and are plotted in Figure 1.

Eighteen selected total RNA samples of this transfer experiment were hybridized onto DNA microarrays. From the resulting microarray data, transcription profiles were analyzed for all of the 253 candidate reference genes. Transcript levels for 16 of these genes exceed the threshold for biological variation of 2.3-fold, and 69 were over the 1.7-

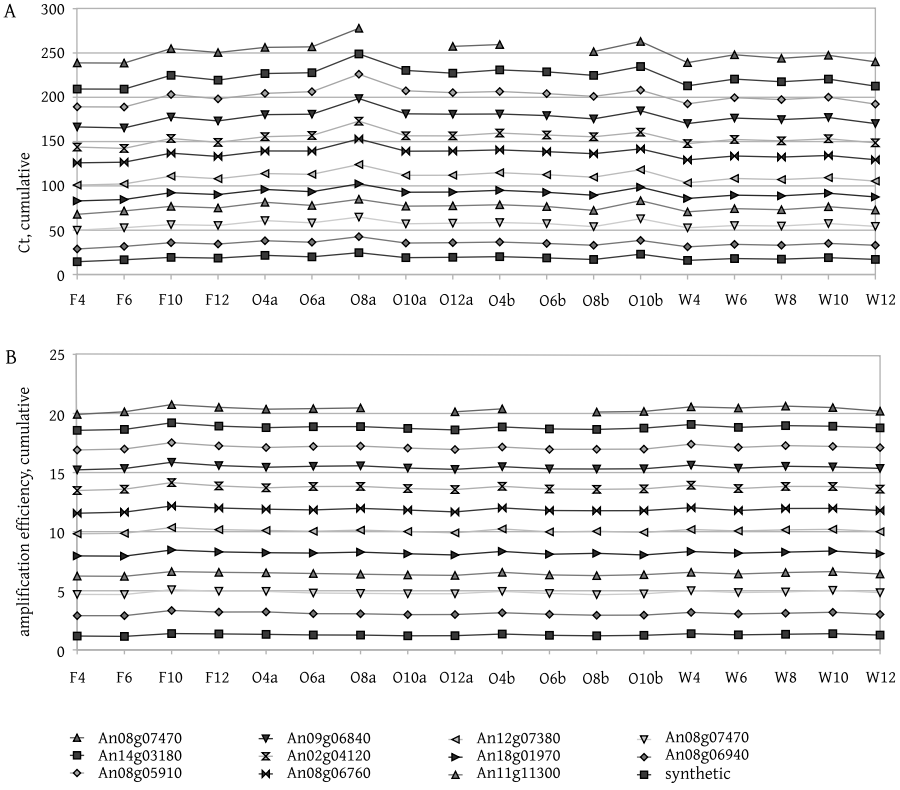


Figure 1 - Cumulative representation of (A) cycle threshold and (B) amplification efficiency values for 11 candidate reference genes examined by qPCR. Samples are represented on X-axis: (F)ructose, (O)live oil, (W)heat diglyceride oil. Olive oil samples are done in duplicate (a,b). Numbers indicate harvesting time in hours after transfer.

fold technical replication threshold. Three of the 11 qPCR-assessed genes (An12g07380, An18g01970, and An14g050) are over 1.7-fold but below 2.0-fold on some arrays.

Synthetic reference gene

An alternative to endogenous reference genes is to spike isolated total RNA with *in vitro* transcribed RNA [87, 238]. The prerequisites for the use of such synthetic transcripts are that (i) the amount of starting material can be determined accurately and reproducibly, (ii) the amplification efficiency for the synthetic gene is not copy number dependent, and that (iii) the efficiency of the reverse transcription reaction is identical for both synthetic transcript and native mRNA.

These prerequisites were examined for our synthetic gene, a kanamycin synthetase encoding transcript. Regarding the starting material, duplicate total RNA measurements deviate no more than 2% in the range of 50 - 500 ng × μL^{-1} (n=36 samples) when measured spectrophotometrically. The copy number dependence for the synthetic transcript was determined using a dilution series up to one-tenth of a million of cDNA amplified from

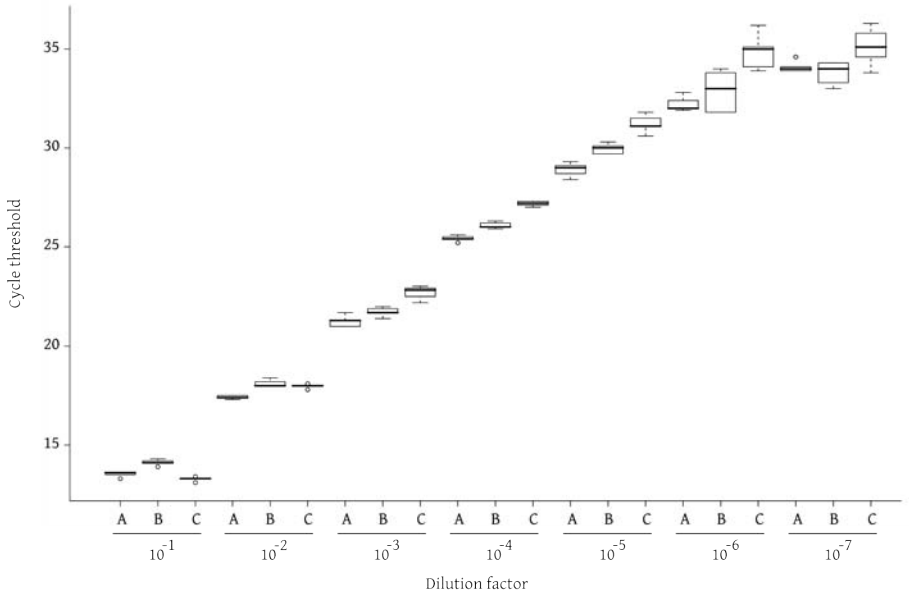


Figure 2 - Average of cycle threshold values of the spike-in synthetic gene measured in a serially diluted cDNA sample. Ct value was measured on three days A, B, and C. Initial spike-in transcript concentration is 1 ng×μL^{-1}. For every day and for every dilution, 5 measurements were taken.

1 ng of synthetic transcript (Figure 2). This dilution series was repeated on three days. The observed cycle threshold values increase slightly during time as is demonstrated in Figure 2. This effect might be due to the effect of freeze-thaw cycles on cDNA quality.

Potential differences in the reverse transcription step between synthetic and endogenous genes were examined as follows. *A. niger* was transferred to shake flasks containing wheat diglycerides as sole carbon and energy source. Mycelium was harvested after 4 hours, and total RNA of two biological duplicates was spiked with 1 ng of synthetic transcript. For each sample, three independent reverse transcription reactions were done. In parallel, an equal mixture of total RNA of both duplicate samples was mixed and processed similarly. A schematic representation of this experimental setup is given in Figure 3. Twenty-seven cDNA preparations were thus obtained. For each cDNA preparation, transcript levels of 3 genes were measured in triplicate: (i) for the synthetic exogenous reference gene, (ii) for the gene encoding isocitrate lyase (An01g09270; isocitrate lyase transcription is expected to be increased because of the presence of oil), and (iii) for gene An14g05050, which has been selected from the list of candidate reference genes of Table 2.

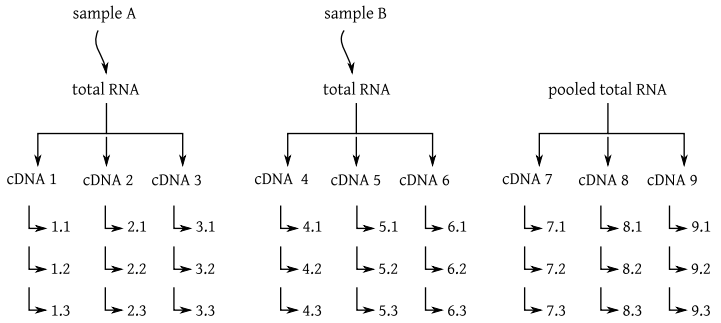


Figure 3 - Experimental design to determine cDNA synthesis variability. Total RNA was isolated from two biological replicates and twice independently reverse transcribed into cDNA. An 1:1 mixture of both total RNA samples was evaluated also. The synthesized cDNA again was split into three parts, and transcript levels were measured independently.

No statistically significant differences were observed between RNA samples for both the synthetic kanamycin synthetase transcript and the endogenous reference gene (t-test, $p = 0.05$). In contrast, a significant difference for the isocitrate lyase gene was found between RNA samples ($p=3.8 \times 10^{-8}$). Although expression of isocitrate lyase is elevated in both cultures compared to non-induced conditions, the mean cycle threshold values of the RNA samples vary slightly: for sample A, the mean Ct value is 16.1, while for sample

B this mean is 16.6. The median Ct value for the pooled RNA is 16.4. This difference shows a shake flask-dependent response of the fungal culture towards the oil. This effect can be accounted for by differences in the amount of fungal cells transferred to the medium. The amplification efficiency remains constant per gene for all three genes examined.

Correlation between technologies

qPCR technology is commonly used to validate DNA microarray results [209], and it has been reported that microarray and qPCR results generally correlate well [178]. We sought to substantiate this correlation by comparison of the qPCR and microarray data for the 11 genes examined in this study. Variation of a reference gene relative to other reference genes can be described by the internal control gene-stability value M as can be calculated by GeNorm [274]. The M-value was calculated for the 11 reference genes using either qPCR or raw microarray data. The genes were ranked by their calculated M-values (Figure 4; the gene with lowest M-value is assigned rank 1, the next-lowest rank 2, and so on).

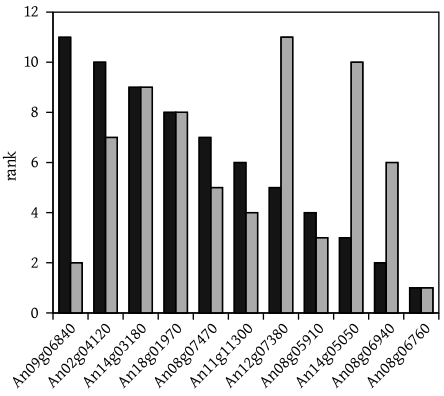


Figure 4 – Endogenous reference genes are ranked by their gene stability measure 'M' as calculated by the GeNorm program. Rank is based on the qPCR data (dark bars) or DNA microarray data (light bars). A lower rank indicates less variability of gene expression.

Four of 11 genes have different rankings between the technologies. Transcript levels for gene An09g06840 vary little on the microarray data, but this gene ranks last based on the qPCR analysis. The reverse observation is made for An08g06940, An14g05050, and An12g07380. Transcript copy numbers do not explain these differences, as the histone An08g06940 is highly expressed on the microarrays with an average probe set signal 4,600 arbitrary units (AU) from a maximum of 25,000, whereas An12g07380 is modestly expressed (average 130 AU). An14g05050 has intermediate average expression levels of 570 AU.

This difference can be explained by the nature of the technologies used. Affymetrix technology makes use of twelve 25-mer probe pairs to cover each gene and the final expression value is composed of fluorescence data for these 12 probe pairs [156]. Furthermore, additional steps – conversion of cDNA to cRNA followed by fragmentation

and hybridization – are needed for the Affymetrix platform. In contrast, only two primers are used for qPCR which amplify a single designated region of the cDNA transcript.

Discussion

Fluctuations in transcript levels provide information on the functional status of the cell under the conditions of interest, and these fluctuations can be analyzed by qPCR. As the actual acquisition of qPCR data is preceded by isolation of total RNA and reverse transcription of mRNA, qPCR data normalization is a necessity to account for differences introduced by all manipulations. How to normalize these data is subject to ongoing debate. One method is the use of stably expressed genes as reference. Reference genes of the Northern blot analysis era were selected by their cellular function or by the availability of genomic sequences, but the expression levels of these genes are known to fluctuate [246, 252]. Software tools have been developed to select the most stable of a set of genes to use for a given tissue or condition [104, 199, 274]. However, often the starting list of genes to select from originates from a limited pool of well-known genes [162, 191, 213]. The availability of microarray expression data for *A. niger* has enabled us to search beyond traditional reference genes following a strategy suggested by Lee and co-workers [152].

We have identified 469 invariable candidate reference genes on microarrays originating from samples cultured under mild conditions. When these genes were either analyzed on stress-imposed conditions or tested on a new experimental condition containing non-preferred carbon sources, the expression of some genes turned out to become variable. The extent of observed variation correlates with the physiological state of the cells: 30% of the listed candidate genes are designated as having variable expression on the stress-imposed arrays, compared to 12% on oil-substrate arrays. This observation can be explained by the introduction of non-biological variation in sample culturing and processing by a different experimenter. However, on six fructose-grown microarrays used in the control growth experiment, only 4 genes are over 2.3-fold, and 33 genes over 1.7-fold differentially expressed. Of these, the gene with highest variability in relative expression is An02g0074. This gene encodes (R)-6-hydroxynicotine oxidase and is involved in nicotine metabolism. The strain used in this experiment is a nicotinic acid auxotroph, which may be a possible explanation for the observed increase.

Recently, ten *A. niger* reference genes for qPCR were selected by Bohle and co-workers based on their role in basic metabolism and by literature screening [28]. The commonly

used reference genes encoding GAPDH and actin were included in their list. We compared the transcript levels of their selected genes with the 11 genes assessed in our qPCR setup that were found by microarray analysis (Figure 5). The reference genes presented in this study show less variation in transcription levels compared to those of the reference genes examined by Bohle and co-workers under both mild and stress-imposed conditions. These differences can be explained by the functional role of the gene products in the fungal cell. Six of 10 genes proposed and tested by Bohle encode enzymes of the glycolysis, the citric acid cycle, or the pentose phosphate pathway. Cox5, a subunit of cytochrome-c oxidase, functions in oxidative phosphorylation. When *A. niger* is transferred to complex or non-preferred carbon sources, or encounters a stressful environment such as carbon starvation or exposure to air, its metabolism will be adjusted drastically. The gene encoding cellular structure-related actin (An15g00560) chosen by Bohle and coworkers is less affected by changes to the environment. This gene is also present in our selection of 253 stably expressed genes. The comparison between both reference gene lists stresses the requirement for reference gene validation for every new experimental condition.

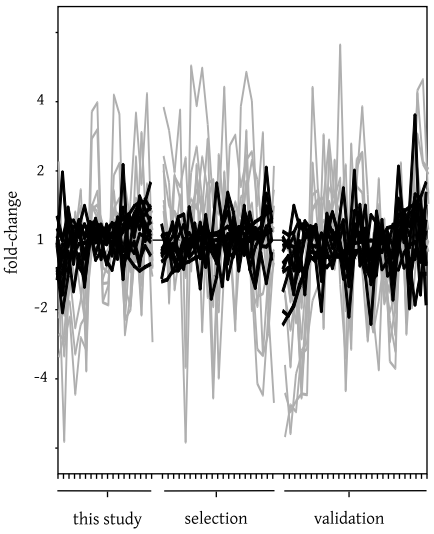


Figure 5 - Normalized intensity of gene expression levels for the 10 reference genes selected by Bohle *et al.* [28] (gray) and the 11 genes examined in this study (black). Microarrays are divided into three groups: this study, 18 microarrays of the experiment described in this study; selection, 21 samples cultured under mild conditions; validation, 27 arrays of stress-imposed cultures.

An alternative to internal reference genes is the use of an external reference gene. In this study we validated a synthetic reference gene encoding a bacterial kanamycin synthase modified by addition of a eukaryotic poly(A)-tail. Transcript stability analysis as measured by the GeNorm algorithm demonstrated that this synthetic transcript is second in stability after the spindle pole body protein-encoding gene, An08g06760. The Affymetrix platform also uses synthetic transcripts: the B2 oligonucleotide control is used to align the array grid; bacterial poly(A)-tailed controls such as *dap* are used to control cDNA synthesis and sample preparation steps; and hybridization controls such as *bioB* are used

to control hybridization [3]. Spike-in experiments are used to validate platform performance [47]. Given this routine use for microarrays it is surprising that synthetic reference genes have only limited use in qPCR based studies [33, 87, 181, 238].

One argument against the use of external reference genes is that quality differences between total RNA samples prior to spiking render differences in transcript levels and may lead to incorrect data interpretation. A second argument is the unreliability of RNA mass determination, or that mass determination ignores situations where the mRNA:rRNA ratio varies between experimental conditions [248]. Our results indicate that spectrophotometric RNA mass determination is robust and reproducible, with less than 2% variation between replicate measurements. The use of microfluidic equipment [179] and integrity qualification methods [230] enables rejection of poor quality total RNA samples. A possible altered total RNA:mRNA ratio will be noticed by an apparent fold change of all endogenous genes relative to the spiked synthetic gene.

Finally, the nature of the synthetic transcript compared to endogenous transcripts was investigated. The synthetic transcript used to spike in this study is of bacterial origin and has no homology to *Aspergillus* sequences. No difference in reverse transcription efficiency was observed for this gene compared to endogenous genes tested.

Conclusions

Current endogenous reference genes are selected by a combination of traditionally selected genes and literature research. With the availability of large-scale transcript data the pool of reference gene candidates can be broadened. Two hundred fifty-three genes were stably expressed under a variety of experimental conditions for *A. niger.* Eleven candidate reference genes were validated by qPCR. The transcript levels of microarray-selected reference genes are less variable compared to other previously published reference genes.

A spike-in reference is shown to be as reliable as an internal reference. No drawbacks assigned to the use of such external spike, such as inability to determine the RNA mass quantity or disturbed cDNA synthesis, were found to be valid. Since a spike-in gene is added after culturing and RNA isolation, its transcript state does not depend on the experimental condition under study. This ability makes its use an excellent tool for tracing effects of the experimental condition on the internal reference genes. The combination of an exogenous reference gene with a number of endogenous genes (which are validated for the experimental condition under study) provides a robust and reliable normalization for qPCR data.

Methods

Culturing conditions

Aspergillus niger NW219 (*cspA1*, *nicA1*, *leuA1*, *pyrA6*), a low-conidiophore mutant derived from N400 (CBS 120.49) was used for experiments conducted for this study. Media are based on minimal medium (MM) [204] and are supplemented with 1.5 mM leucine, 8 µM nicotinic acid amide and 5 mM uridine. Mycelium was precultured (30 °C, 225 rpm) in supplemented MM containing 100 mM D-fructose as carbon source as well as 0.1% yeast extract and 0.1% casamino acids using 1×10^6 spores × mL^{-1} as inoculum. After 18 hours, the mycelium was recovered by filtration over a Büchner funnel and was washed twice with MM. Five ± 0.2 grams wet biomass were transferred to new shake flasks that contained as carbon source either 20 mM D-fructose, 0.5% (v/v) refined olive oil (Sigma), or 0.5% (w/v) of a diglycerides-enriched wheat fraction obtained as described in [56]. 1% (w/v) Gum Arabic (Sigma) was used as emulsifier. Mycelium was harvested after 4, 6, 8, 10 or 12 hours after transfer, filtrated, and snap-frozen in liquid nitrogen. All carbon source and time-point combinations were performed in duplicate. Duplicate samples were cultured in different weeks. The 4, 6, and 8 hour time point duplicate samples for the three carbon sources were hybridized onto in total 18 DNA microarrays.

RNA isolation and cDNA synthesis

Total RNA was isolated from ground mycelium using TRIzol Reagent (Invitrogen) followed by chloroform extraction according to the manufacturers recommendations.

Isolated total RNA was dissolved in 10 mM Tris-HCl pH 8.0, heated for 2 min at 70 °C, and diluted in two steps to a concentration of 200 ± 5 ng × μL^{-1} (NanoDrop ND-1000 Spectrophotometer, NanoDrop Technologies). 1.00 µg of total RNA was spiked with 0.1 ng of a synthetic RNA transcript encoding kanamycin synthetase (Promega) followed by DNaseI-treatment. cDNA was synthesized using the Omniscript reverse transcriptase kit (Qiagen). Four units of reverse transcriptase enzyme were added to the DNase-treated total RNA in a final volume of 20 µL. cDNA synthesis reaction was at 37 °C and subsequently water was added to a final volume of 400 µL. Diluted cDNA was stored at -20 °C until use.

Quantitative real-time RT-PCR (qPCR)

Oligonucleotide primers were designed using Primer3 software [218]. Primer criteria were: length 20 to 25 nucleotides, melting temperature 60 ± 1 °C , GC contents 45-55%. Amplicon product sizes range between 120 and 200 nucleotides and had a melting temperature of 80 ± 5 °C. Primer pairs were designed for the 3'-end of the ORF and covered an intron-spanning region when possible. For qPCR measurements, 25% (v/v) of diluted cDNA was added to 2× SYBR PCR mastermix (Abgene) and PCR primers using a CAS-1200 pipetting robot (Corbett Life Science). Final concentration for each primer was

1.4 µM. A Rotorgene-3000 qPCR machine (Corbett Life Science) was used to run the following program: 15 min at 95 °C; followed by 40 cycles of 15 seconds at 95 °C, 15 seconds at 58 °C and 20 seconds at 72 °C. The program ended with a melting procedure. Determination of Ct-values and amplification efficiencies was done with the "comparative quantitation" method of the Rotorgene software (version 6.0.19).

DNA microarrays

Several DNA microarray datasets used for selection were available in our laboratory. One study of 30 microarrays is described elsewhere [173]. In short, *A. niger* N400 grown in shakeflask was transferred to media containing various sugars as sole carbon source, and harvested up to 8 hours or after 24 hours after transfer. In other studies from our laboratory, in total 20 microarrays were hybridized with mycelium harvested under starvation conditions or derived from fermentor cultures using more complex carbohydrate carbon sources (*e.g.* sugar beet pectin) [H. Kools and P. Schaap, unpublished data].

For the microarrays hybridized in this study, 5 µg of isolated total RNA was processed following the Affymetrix protocols for eukaryotic target preparation [3]. Biotin-labeled cRNA was synthesized using the BioArray High Yield RNA transcript labeling kit (Enzo Life Sciences). Quality of both total RNA and cRNA were analyzed on an RNA6000 Nano Assay Labchip (Caliper Technologies) on an Agilent 2100 Bioanalyzer (Agilent). Fifteen µg of labeled cRNA was used for hybridization following the Affymetrix protocol for eukaryotic target hybridization and staining. Microarrays were scanned on an Agilent G2500A GeneArray scanner. Raw intensity values and "present" or "absent" flags were extracted using MAS5 software (Affymetrix) and analyzed in GeneSpring GX 7.2 (Agilent technologies).

Candidate reference gene selection

Probe sets are defined reliable when flagged by Affymetrix detection significance call as "present" and have a minimum raw value of 100 AU on all microarrays used for candidate gene selection and validation in this study. Of 14,554 probe sets, 2,429 probe sets including 15 Affymetrix control probe sets fulfilled this criterion. Variable genes were selected based on their standard deviation following Lee *et al.* [152], removing all probe sets with values higher than the median standard deviation for the control probes (mean SD = 0.20). Fold change criteria were derived from a study investigating the variance components for controlled *A. niger* fermentor experiments [this thesis, Chapter 3]. Less than 0.1% of all probe sets is over 1.7-fold differentially expressed between technical replicates of one sample onto two independent arrays. Between biological replicates 0.5% of all probe sets is over 2.3-fold changed.

Acknowledgments

The authors would like to thank S. Basmagi and P. Schaap for access to the in-house fungal COG database, and DSM Food Specialties for access to the DSM *A. niger* Affymetrix microarrays.

Chapter 3

Variance components analysis reveals contribution of sample processing to transcript variation

Douwe van der Veen [‡], José Miguel Oliveira [‡], Willy A.M. van den Berg, and Leo H. de Graaff

[‡] These authors contributed equally to this work

Appl. Environ. Microbiol. 2009. **75**:2414-2422

Abstract

Proper design of DNA microarray experiments requires knowledge of biological and technical variation of the studied biological model. For the filamentous fungus *Aspergillus niger*, a fast, quantitative real-time PCR based hierarchical experimental design was used to determine this variation. Variance components analysis determined the contribution of each processing step to total variation: 68% is due to differences in day-to-day handling and processing, while the fermentor vessel, cDNA synthesis, and qPCR measurement each contributed equally to the remainder of variation.

The global transcriptional response to D-xylose was analyzed using Affymetrix microarrays. Twenty-four statistically differentially expressed genes were identified. These encode enzymes required to degrade and metabolize D-xylose containing polysaccharides, as well as complementary enzymes to metabolize complex polymers likely present in the vincinity of D-xylose containing substrates. These results confirm previous findings that the D-xylose signal is interpreted by the fungus as the availability of a multitude of complex polysaccharides.

Measurement of a limited number of transcripts in a defined experimental setup followed by variance components analysis is a fast and reliable method to determine biological and technical variation present in qPCR and microarray studies. This approach provides important parameters for experimental design of batch-grown filamentous cultures, and facilitates the evaluation and interpretation of microarray data.

Introduction

Culturing filamentous organisms such as *Aspergillus niger* is difficult to reproduce when compared to culturing unicellular organisms. Filamentous growth is characterized by elongation and branching of hyphae, cylindrical cells that increase in length by growth at one end. *De novo* biosynthesis and active enzyme production occurs mainly at the hyphal tips. In regions at distance from the tip, the hyphae age and become biologically less active [291]. This hyphal growth is the result of adaption to the habitat of the organism which enables it to spread over and penetrate surfaces and cross over nutrient-depleted gaps [42]. However, under laboratory conditions, attachment of fungal mycelium to fermentor baffles and other extremities introduces heterogeneous growth that only to a certain extent can be suppressed (for instance, by cooling the fermentor headplate). The growing mycelium increases the culture broth viscosity, which reduces the mass transport of nutrients, oxygen, and heat, and affects the mixing characteristics in a fermentor or shakeflask over time. Physical agitation and shear stress can cause uncontrolled breakage and fragmentation of the mycelia [285].

Recent technologies such as global transcriptome analysis by DNA microarrays or quantitative real-time PCR (qPCR) require the use of replicate biological samples for high quality data. Given the difficulties in culturing *A. niger*, to obtain transcript data without waste of resources requires proper experimental design. Key in designing such experiments is to determine how much replication is needed – the sample size [284]. A larger amount of replicates leads to increased statistical accuracy of measurement, whereas insufficient replication impedes data analysis. The required number of independent samples depends on a variety of factors: the organism under study and its biological variation, the magnitude of expected gene response, the statistical power to detect the genuine gene response to a condition, and the false discovery rate [123, 284].

Consensus is emerging on what a ‘proper microarray experiment’ comprises [123]. According to this consensus, defining the experimental objectives and requirements is a necessity before actually starting an experiment [292]. During the experiment, standardized protocols and methods reduce the variability that is introduced in each process step [221]. A selected experimental design makes data analysis and interpretation as simple and as powerful as possible. Finally, for publication of the microarray results and data, the 'Minimal Information About a Microarray Experiment' (MIAME) guidelines [35] are adopted. Notwithstanding this consensus, Jafari and Azuaje have published an extensive review on papers describing microarray methods and applications of microarray technology [125], and concluded that recently published gene expression data analysis studies often lack key information required to interpret and evaluate published data.

The goal of the present study was to minimize the variation in *A. niger* batch fermentations by optimization of protocols and procedures. The variation between fermentations was determined by a variance components analysis of data obtained by qPCR. This relatively inexpensive technology is used to measure transcript levels for few genes in many samples simultaneously. Furthermore, qPCR is routinely used to validate microarray results [74, 178].

The effects of the optimization and quality control measures for our experimental set-up were assessed by examination of the global transcriptional response towards induction with D-xylose. The xylanolytic system of *A. niger* is under control of the transcriptional activator XlnR, and genes under its control are well-documented [271]. Recently, the transcriptional response towards D-xylose was examined by microarray analysis for three *Aspergillus* species [11]. The availability of these data on the transcriptional response towards D-xylose allows for validation of the biological response observed during our studies.

Results

Selection of conditions of the response system

The xylanolytic system of *A. niger* consists of enzymes that degrade cellulose and hemicellulose, is under control of the transcriptional regulator XlnR [271], and can be induced by the monosaccharide D-xylose. XlnR-controlled genes are subject to carbon catabolite repression by D-xylose [59]. To reduce this repressing effect, the D-xylose concentration that induced the xylanolytic system but had the least repressing effect on XlnR-controlled genes was determined prior to the variance components analysis experiment. Three fermentor cultures were induced with either 0.1, 1, or 50 mM D-xylose, and sampled for up to 4 hours. Expression levels for two XlnR-controlled genes, *xlnB* and *xlnD*, were followed by qPCR. These genes encode endo-xylanase B [139] and β-xylosidase [3269], respectively.

The observed transcript levels for *xlnB* and *xlnD* are given in Figure 1. A D-xylose concentration of 0.1 mM is able to induce both genes, with an expression ratio for *xlnD* of 190-fold at 60 minutes after induction. This lowest concentration of 0.1 mM D-xylose also discriminates best between non-induced and induced states: for this concentration, transcript levels increase until 60 minutes and return to pre-induced levels in the next hour. For both the 1 mM and 50 mM concentrations, elevated transcript levels for *xlnB* and *xlnD* could be detected up to three hours after induction. We decided to induce fermentor cultures with 0.1 mM D-xylose and to sample 1 hour after induction.

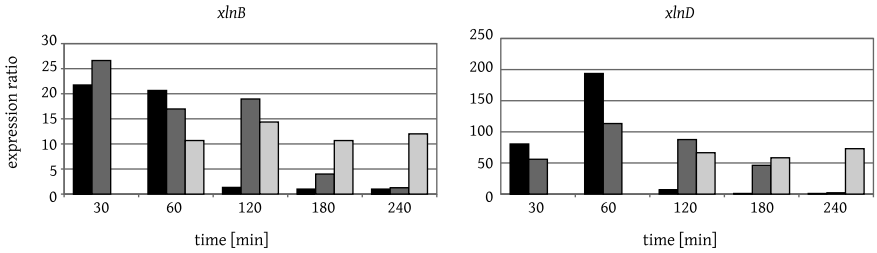


Figure 1 - The expression ratio of the *xlnB* and *xlnD* genes measured after induction with either 0.1 (black), 1 (dark-gray), or 50 (light-gray) mM of D-xylose. The sampling time in minutes is presented on the horizontal axis. No transcript levels were determined for the 50 mM fermentations for both genes at 30 minutes. At 60 minutes, a bad qPCR run prevented *xlnD* ratio calculation.

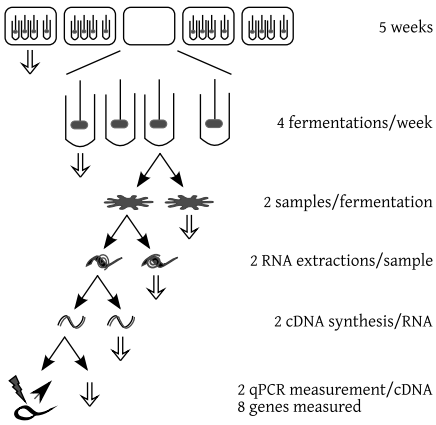


Figure 2 (left) - Schematic representation of the hierarchical experimental design. In a 5 week period, 4 fermentors were run of which three were induced with 0.1 mM D-xylose and one with 0.1 mM sorbitol. Each fermentation was sampled twice; total RNA was extracted twice independently from each mycelial sample; this total RNA was split in two and independent cDNA synthesis reactions were performed. Each cDNA sample was used to measure 8 genes using qPCR, each measurement in duplicate. Gray arrow indicate an identical processing step.

Experimental design

The aim of this experiment was both to minimize all variation introduced in each of the steps preceding a qPCR measurement or microarray scan, and to determine their magnitude. After optimization of the processing protocols, we investigated the variation in our experimental set-up quantitatively by qPCR. For this, a hierarchical factorial experiment was designed (Figure 2). We performed one fermentation cycle per week for five weeks, using 3 identical fermentations per week. These cultures were induced with 0.1 mM of D-xylose. A fourth culture was randomly selected as 'non-induced control'. Each processing step following induction was executed in duplicate, starting with the independent harvesting of mycelial samples. This led to 128 individual qPCR measurements for a single fermentation, or a total of 1,920 individual measurements.

Another 768 measurements originated both from non-induced cultures and from culture samples taken before induction. In total, 2,688 qPCR measurements were made.

In the elongation phase of a qPCR reaction, SYBR-green dye molecules intercalate with double stranded DNA that is formed during the reaction. The dye-DNA complex absorbs blue light and emits green light [259], which is measured and plotted against the cycle run number. The resulting sigmoid-shaped curve is used to extract the cycle threshold and amplification efficiency values. The cycle threshold value represents the amount of transcripts for a specific gene in the cDNA template upon start of the PCR reaction, with lower values representing more copies of a transcript present in the cDNA pool. The amplification efficiency is a primer-pair specific value that represents the efficiency of the PCR reaction, and hence is independent from the experimental conditions apart from the qPCR run itself [253]. Insight into the variation that is introduced in all steps except the day-to-day and fermentor-to-fermentor variation is deduced from the expression ratio. This ratio expresses the relative change of a gene's transcription level compared to a non-induced sample. A ratio greater than 1 indicates elevated transcription of a gene, while a ratio of 1 indicates no change. Table 1 gives descriptive statistics of the data set. Variance components were estimated for each of the 3 variables, and were subdivided by the 8 genes followed in this study (Table 2).

	Cycle threshold		Amplification		Expression ratio	
Gene	Samples	average ± SD	Samples	average ± SD	Samples	average ± SD
An08g05910	230	24.08 ± 1.22	230	1.70 ± 0.026	226	1.06 ± 0.43
An02g04120	232	21.40 ± 0.98	232	1.72 ± 0.044	228	1.05 ± 0.43
An14g05050	236	25.69 ± 1.47	236	1.68 ± 0.032	232	1.02 ± 0.42
An08g06940	236	16.49 ± 0.57	236	1.70 ± 0.024	232	0.88 ± 0.32
An15g01860	233	24.83 ± 1.43	233	1.71 ± 0.015	229	2.45 ± 1.74
kanamycin	234	16.08 ± 0.49	234	1.74 ± 0.028	234	1.00
xlnB	236	22.62 ± 1.38	236	1.73 ± 0.018	232	16.42 ± 9.35
xlnD	235	16.08 ± 0.82	235	1.73 ± 0.027	231	137.73 ± 94.90

Table 1 - Descriptive statistics for the two by qPCR measured and the calculated values used for cycle threshold, amplification efficiency, and expression ratio. The total number of measurements per gene is 240. qPCR runs that did not meet quality control standards account for differences between theoretical and actual measured samples.

Gene	Week	Fermentor	Biomass	RNA Sample	cDNA Sample	qPCR
	Cycle threshold [%]					
An08g05910	78.8	5.9	1.3	2.5	4.2	7.3
An02g04120	66.2	1.9	0.0	0.3	11.5	20.1
An14g05050	81.4	3.0	2.7	3.9	3.2	5.9
An08g06940	68.1	5.5	1.0	14.1	8.3	2.9
An15g01860	29.6	60.4	0.6	2.8	3.1	3.6
kanamycin	67.5	0.0	0.0	18.1	10.1	4.3
xlnB	69.1	10.7	0.0	4.7	9.3	6.2
xlnD	4.8	19.4	2.9	47.7	8.2	17.0
	Amplification efficiency [%]					
An08g05910	68.9	2.1	2.8	0.0	1.1	25.1
An02g04120	75.9	3.1	0.0	0.0	2.2	18.7
An14g05050	74.7	0.1	0.0	0.0	0.3	24.9
An08g06940	77.8	0.1	0.0	0.0	4.6	17.5
An15g01860	30.1	5.3	0.0	1.5	8.2	54.8
kanamycin	71.5	0.8	0.0	7.5	2.1	18.1
xlnB	31.2	6.2	7.1	0.0	11.3	44.3
xlnD	32.6	25.2	1.9	9.1	6.6	24.7
	Expression ratio [%]					
An08g05910	0.0	36.4	10.7	0.0	24.6	28.3
An02g04120	2.1	28.9	0.0	4.3	21.5	43.2
An14g05050	0.0	30.6	9.8	2.8	24.5	32.2
An08g06940	0.3	58.5	2.3	12.7	19.8	6.3
An15g01860	6.0	75.9	0.5	6.8	1.9	9.0
kanamycin	0.0	0.0	0.0	0.0	0.0	0.0
xlnB	11.7	62.0	0.0	6.0	9.4	10.9
xlnD	0.0	80.7	1.6	3.6	11.5	2.6

Table 2 - Estimates of the relative variance compared to total variance (which is set at 100%). Columns indicate the terms of the random model used to describe variance and correspond to the experimental procedure outlined in Figure 2.

DNA microarray analysis results

To investigate the differences in global transcript levels between fermentations done in this study, six fermentor samples were selected for hybridization onto DNA microarrays. These samples were selected based on differences in their distribution over the weeks and fermentor vessels used, their biomass density, and in the *xlnD* expression levels observed in these samples by qPCR (Table 3).

To analyze technical variation that is introduced within microarray processing steps, total RNA of one mycelial sample was split (Table 3, C and D), processed independently, and hybridized onto two microarrays. RMA signal values were calculated and plotted (Figure 3). Only the RMA signal values for 5 of 14,554 probe sets differ over 1.5-fold but not more than 2.1-fold between the technical replicates. No mRNA sequences hybridize

Sample	Carbon source	Week	Fermentor	Biomass	RNA sample	*xlnD* ratio ‡	Dry wt [g/liter]
A	D-xylose	2	3	2	1	20	1.84
B		3	2	2	1	300	1.41
C		5	3	1	1	50	2.20
D		5	3	1	1	50	2.20
E		4	5	2	1	120	1.93
F	sorbitol	5	5	1	1	1	1.64
G		3	4	1	1	1	1.92

Table 3 - Properties of the RNA samples selected for DNA microarray hybridization. Coding corresponds to the by qPCR measured data, e.g. 'biomass 1' refers to the first of two independent mycelial samples taken from a fermentation. ‡ The *xlnD* ratio is the average of 4 expression ratios calculated from the 4 replicate qPCR measurements.

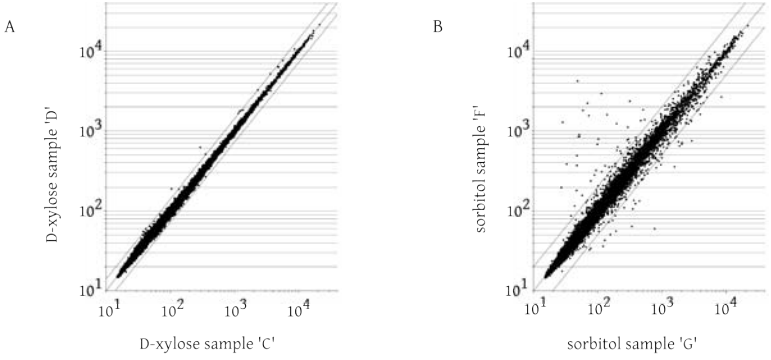


Figure 3 - Technical variation between samples. Scatter plot of microarray data. Panel A: Two RNA samples derived from one mycelial sample (Table 4, C and D) were processed independently and hybridized to two Affymetrix microarrays. The RMA-normalized data of all 14,554 probe sets are plotted in a scatter plot. Panel B: total RNA from two independent non-induced control fermentations (sorbitol; Table 4, F and G), are hybridized and RMA-normalized data are plotted. The lines above and under the diagonal indicate 2-fold difference relative to the diagonal.

A. niger	Secretion signal	Description	Motif
An01g00780	yes	*xlnB/xynB* :: xylanase-B	-216
An01g09960	yes	*xlnD* :: β-xylosidase	(-250) -147
An14g05800	yes	*aguA* :: α-glucuronidase	-277, (-183)
An09g00120	yes	*faeA* :: ferulic acid esterase A	-265, (-121)
An18g03570	yes	*bglA* :: β-glucosidase	
An14g02760	yes	*eglA* :: endoglucanase A	
An09g03300	yes	glycosyl hydrolase, family 31	-138
An03g00500		glycosyl hydrolase, family 30	
An15g01500		sugar transporter	417
An11g01100		sugar transporter	
An03g01620	yes	sugar transporter	-407
An06g00560		sugar transporter	-415
An01g10920		*ladA* :: L-arabitol dehydrogenase	-334
An01g03740		*xyrA* :: D-xylose reductase	311
An12g00030		*xdhA* :: xylitol dehydrogenase	-333
An07g03140		*xkiA* :: D-xylulokinase	(-672), -482
An02g03590		galactose-1-phosphate uridylyltransferase (similar to *S. cerevisiae* GAL7p)	371
An11g10890		UDP-glucose 4-epimerase (similar to *S.cerevisiae* GAL10p)	577
An07g03160		*talB* :: transaldolase	-195
An08g01740		aldehyde reductase	-765
An02g13980	yes	conserved hypothetical protein (mono-oxygenase?) ; cytochrome P450 domain	-540, (+26)
An06g00830		weak similarity to hypothetical transcription regulatory protein	
An01g11180		hypothetical protein, FAD/FMN containing dehydrogenase	329
An11g05340	yes	fructosamine oxidase	

Table 4 - Differentially expressed genes after induction with D-xylose. All 24 genes are upregulated upon D-xylose induction. The location of the *cis*-acting XlnR motif, 5'-GGNTAAA-3', is indicated in base pairs counted upstream of the ATG start codon. Values in parentheses indicate the reverse direction of this motif.

to these probe sets: 28S ribosomal RNA hybridizes to four of five probe sets while the fifth probe set hybridizes to one of the control transcripts used in the Affymetrix array platform, the 5' region of the *E. coli bioD* transcript. This result indicates that the specific microarray handling and processing steps contribute little to overall variation.

A transcriptome analysis revealed 24 genes that are significantly differentially expressed in D-xylose-induced cultures compared to sorbitol-grown cultures (Table 4). This gene list contains both *xlnB* and *xlnD*, the two genes used in the qPCR measurements to assess the induction of the xylanolytic system, as well as a further 8 enzymes that are known to be under XlnR control [271]. The presence of four sugar transporter-encoding genes, as well as two genes encoding glycosyl hydrolases, is in accordance with the view that D-xylose initializes a response to degrade complex carbohydrates such as hemicellulose. In addition, two genes that encode enzymes of the Leloir pathway, the classical pathway of D-galactose catabolism, are upregulated in the induced samples. An02g03590, the ortholog of the *Saccharomyces cerevisiae* GAL7 protein, encodes the second enzyme of the Leloir pathway, while the required glucose epimerase activity is encoded by An11g10890, the ortholog of *S. cerevisiae* GAL10. The first step of this pathway, the galactokinase that phosphorylates D-galactose to D-galactose 1-phosphate, is encoded by An16g04160. This gene is about 2-fold induced by D-xylose, but a p-value of 0.604 excludes this gene from our list of statistically differentially expressed genes. Finally, for six genes in our list of differentially expressed genes, their roles in the cellular response towards D-xylose is difficult to assess.

Discussion

Experimental variation

Although, in theory, a protocol is a rigid and established code that describes all proceedings relating to an experiment, in practice, protocol steps can be subject to interpretation by different experimentors. A precise description of how to handle given steps and removal of unclear or ambiguous sentences decreases the necessity to interpret the exact meaning of a protocol phrase, which in turn leads to more reproducible experiments. The effects of our optimization and quantitative determination of the variation in our laboratory set-up was determined by qPCR, a cost-effective yet powerful technique to study transcript abundance.

Important parameters for high-quality qPCR measurements are the specificity of a primer pair for its target, the amplification efficiency, and reproducibility [41, 161]. Typically, when a newly designed primer pair is tested prior to use in quantification, a

range of different annealing temperatures or primer concentrations is examined to identify conditions where only a single DNA fragment is amplified. In this study, a multitude of target genes were measured using a single PCR profile and primer concentration. The possibility that this approach affects the robustness of our qPCR measurements was assessed by the examination of the variation that is introduced in the amplification efficiency values. Within-gene variation is minimal, even when the amplification efficiency per gene varies. The most variable amplification efficiency values are for gene An02g04120, with a mean value of 1.72 and standard deviation of 0.04 (n=326). The reproducibility of this generalized method can be derived as well from the variance analysis while looking at the the variance component 'By qPCR measurement', which includes pipetting and the actual measurement. The median relative contribution is 6.1% of the total variation (Table 3). Transcript An02g04120 again shows the highest variation, 20.1%, in this step. When this percentage is placed in its biological context, 95% of all cycle threshold values will range between 21.3 ± 0.9, which translates to 1.6-fold differences. We conclude that our standardized quantitative PCR method is both precise and accurate.

Before the actual determination of the sources of variation in our experiment we hypothesized that most variation is introduced by differences in the day-to-day handling and processing. The next-largest variation was anticipated to be the use of the individual fermentor vessels. The results of the variance components analysis comply with our initial hypothesis: day-to-day variation contributes to about 70% of the total variation (Table 2). This step includes the growth and harvesting of spores, the preparation of fermentation media, and the assembly of fermentors. The large effect that the day-to-day variation has on the total variation can effectively be excluded by examination of the expression ratio. This ratio presents the relative change in a gene's transcript level between pre-induced and post-induced fermentation conditions. As this ratio is calculated from expression data of samples taken from the same fermentor, effectively this result cancels out the contribution of day-to-day variation. When fermentor cultures are compared by this gene expression ratio, the variance components analysis of the ratio-derived data shows that the three steps of fermentation, cDNA synthesis, and actual qPCR measurement each contribute about equally in case of the four endogenous reference genes (Table 2).

The effect that the individual fermentation vessels have on the variation in transcript level is 60% and 80%, respectively, for the two D-xylose induced genes, *xlnB* and *xlnD*. For the malate synthase encoding gene, the effect is 75%. Different amounts of fungal cells present in a fermentor cannot explain this effect as only a weak correlation of cycle threshold values with the biomass concentration is found (*i.e.* for *xlnB*, $R^2 = 0.35$). One explanation is that small differences between fermentor headplates and vessels result in

unique mixture characteristics in each fermentor. These fermentor-specific effects are reflected in small physiological differences between cultures which may account for the observed differences in transcript levels. As samples are taken 1 hour after induction with D-xylose, such differences may affect the actual D-xylose concentration per fermentor, leading to a high reproducibility of a gene's expression within a fermentor but variation of its expression between fermentors. For example, *xlnD* induction is on average 240-fold for fermentors with headplate '2', but only 100-fold for fermentors with headplate '4'.

D-xylose induced genes assessed by microarray analysis

The qPCR measurements clearly showed elevated transcript levels of *xlnB* and *xlnD* after induction with D-xylose. For *xlnB*, the average increase over all 15 fermentations was by 16-fold compared to non-induced conditions, while for *xlnD* the average increase was by 138-fold. For the 6 DNA microarrays analyzed, the average increase is by 6-fold for *xlnB* and by 110-fold for *xlnD* after induction with D-xylose. The results for a gene measured by both methods are in good agreement: for instance, the Spearman correlation coefficient for *xlnD* measurements obtained by qPCR and microarray technologies is 0.90.

Comparison of the non-induced and D-xylose induced samples that were hybridized onto microarrays revealed twenty-four genes that are statistically differentially expressed after induction with D-xylose (Table 5). Degradation of complex polysaccharide substrates starts with the uptake of signal molecules such as D-xylose, that activate specific induction pathways resulting in the expression and secretion of enzymes necessary to degrade and metabolize the polysaccharide. However, as D-xylose rarely is found by itself under natural conditions, it is likely that *A. niger* interprets the presence of D-xylose as proxy for the availability of complex carbohydrate polymers, such as (hemi)-cellulose. Not only is this heterogenic response reflected in the induction of secreted enzymes, but also in the activation of multiple metabolic pathways (Figure 4). For instance, genes encoding the second step in L-arabinose metabolism and enzymes of the classical D-galactose catabolic route are as well significantly upregulated.

Five genes appear to encode enzymes that might well be related to (hemi)-cellulose degradation, but of which the exact function has not been elucidated. An08g01740 encodes an aldehyde reductase. Reduction of the carboxylgroup of carbohydrates by aldehyde reductases, such as the reduction of D-xylose to xylitol, is a first step in the catabolism of many monosaccharides, and the product of this gene might well play a role in the catabolism of monosaccharide substrates that are the product of (hemi)-cellulose degradation. Products of D-xylose or L-arabinose catabolism are channeled into

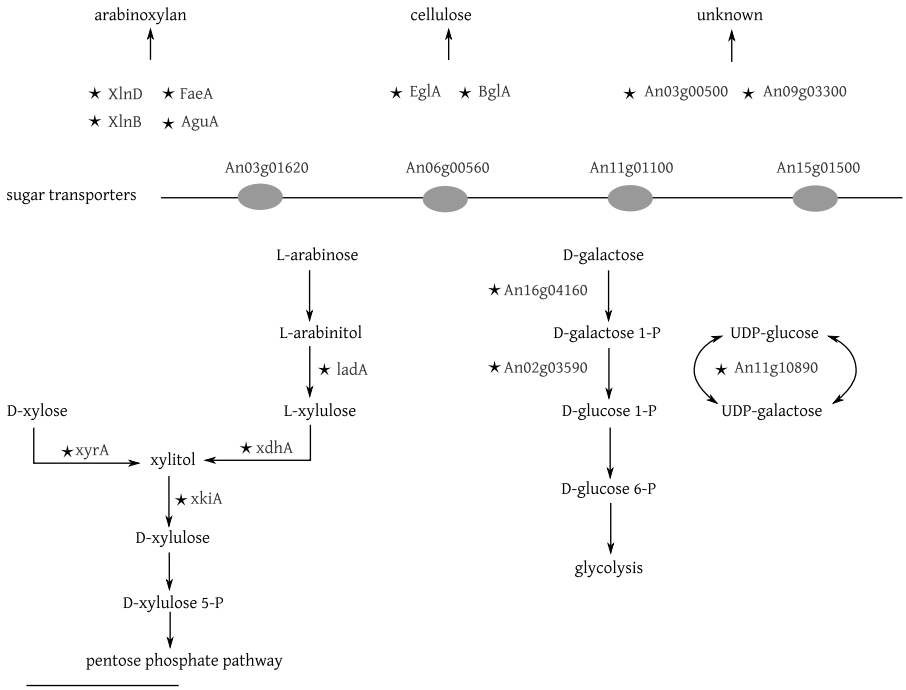


Figure 4 - Schematic representation of the enzymes encoded by the genes that are significantly differentially expressed after induction by D-xylose. Top: extracellular enzymes acting on the complex polysaccharides arabinoxylan and cellulose. Middle: sugar transporters. Bottom: metabolic routes of the degradation of D-xylose, L-arabinose, and D-galactose. The enzymes encoded by significantly differentially expressed genes are indicated by stars.

the non-oxidative branch of the pentose phosphate pathway. This pathway in turn is linked to glycolysis by the enzymes transaldolase and transketolase [30]. While the *talB* transaldolase is 6-fold upregulated in D-xylose induced samples, both transketolase-encoding genes as well as the major transaldolase encoding gene of *A. niger* have no differential expression levels under induced and non-induced conditions.

The transcriptional regulator XlnR activates the transcription of (hemi)-cellulose degrading enzymes [271] as well as the transcription of genes encoding metabolic enzymes [271, 275]. A *cis*-acting element in the promoter region of D-xylose-induced genes, described as 5'-GGCTAAA-3' [271] and proposed to be generalized to 5'-GGNTAAA-3' after comparative transcriptome analysis of three *Aspergillus* species [11], was detected in all but 5 statistically significant differentially expressed genes (Table 5), and also, the GAL1p ortholog An16g04160 has this consensus motif present at 153 base pairs upstream in its promoter region.

In this and other studies [11] it is observed that non-strictly D-xylose metabolism-related genes have elevated transcript levels upon D-xylose induction. The most likely explanation is that the fungus fine-tunes its response towards the diversity of complex substrates it encounters by co-ordinated action of party overlapping regulatory systems. Expression of the XlnR-controlled ferulic acid esterase A is greater when *A. niger* is induced with both D-xylose and ferulic acid relative to induction by D-xylose or ferulic acid alone [58]. Aromatic compounds such as ferulic acid not only are part of arabinoxylan, but also of pectin. No XlnR related motif is present in the promoter region of the polysaccharide-degrading enzymes acting on cellulose, BglA and EglA, nor in the promoter region of one of the uncharacterized family 30 glycosyl hydrolases. The induction of these genes without the presence of the consensus motif suggests the action of other transcriptional activators besides the xylanolytic activator XlnR. Interestingly, one of the genes of unclear function encodes a hypothetical transcription regulatory protein, and its expression pattern correlates strongly with the three genes of the Leloir pathway.

Andersen and co-workers have described a conserved set of twenty-three genes for which transcription is elevated on D-xylose medium [11]. Nine of these D-xylose responsive genes are not present in our gene list: three genes encoding sugar transporters, three glycosyl hydrolases, and three metabolic enzyme-encoding genes (encoding an aldose 1-epimerase, a short-chain dehydrogenase, and an aldehyde reductase). The difference between the two gene lists can be explained by the experimental approach chosen: Andersen and co-workers have grown *A. niger* in either D-glucose or D-xylose as sole carbon source, and have sampled at D-xylose levels around 12 mM. In this study, sorbitol-grown cultures were induced with minute concentrations of D-xylose only.

In conclusion, the work presented in this study has resulted in an improved method for the generation of high-quality qPCR and microarray data from fermentations of the filamentous fungus, *Aspergillus niger*. The decreased variation improves data quality and eases data analysis, which is an essential prerequisite to study transcript profiling and gene regulation.

This work provides new insights into the mechanisms following D-xylose induction. The data link xylose metabolism not only to L-arabinose metabolism but also to D-galactose metabolism.

Materials and Methods

Strain and spore preparation

A. niger 872.11 (Δ*argB*, *pyrA6*, *prtF28*, *goxC17*, *cspA1*) is derived from CBS 120.49. All media were based on minimal medium (MM) [204], contained 100 mM sorbitol as carbon source and were supplemented with uridine and arginine. To obtain spores, 20 spores per mm^2 were plated onto Complete Medium (CM) plates [204], incubated 5 days at 30 °C, and allowed to mature at 4 °C for 24 hours. The spore suspension was washed and stored at 4 °C until use.

Fermentation

Four glass 2.5-liter fermentors (Applikon) with 2.2 Liters of MM were kept at a constant temperature of 30 ± 0.5 °C while fermentor headplates were kept at 8 °C. 1.0×10^6 spores per mL were added to the fermentor. During germination, each fermentor was aerated through the headspace (50 L/h) and stirred at 300 rpm. When dissolved oxygen levels dropped below 60% for over 5 minutes, the stirrer speed was set to 750 rpm and aeration was switched to sparger inlet. This switching time point was defined as T=0 hours. Each fermentor was induced with either 0.1 mM sorbitol or D-xylose at T=14 hours. Samples were snap-frozen in liquid nitrogen directly after filtration, with less than 20 seconds between sampling and snap-freezing.

RNA isolation

Frozen mycelium was ground for 40 seconds using a dismembrator (Braun-Melsungen). A trizol-chloroform extraction preceded total RNA extraction with RNeasy mini columns (Qiagen) according to the manufacturer's protocol for yeast. RNA integrity was assessed on an Experion system (Biorad) by visual inspection of the electropherograms. Graphs depicting RNA integrity categories were used as visual aid [230]. Electropherograms approximating a RIN-number of 8 or lower, or with 28S/18S ratio below 1.8, were discarded.

qPCR measurements

Variation in transcript levels was determined for 7 *A. niger* genes and a synthetic control RNA transcript (Table 5). This synthetic control RNA transcript, a bacterial kanamycin synthetase-encoding gene fused to a eukaryotic poly-(A) tail (Promega), is spiked to total RNA prior to cDNA synthesis and can correct for varying efficiencies of the reverse transcription or the PCR reaction itself [114]. The first four genes of Table 1 were used as endogenous reference genes. These *A. niger* reference genes showed little variation in transcript levels on over 60 microarrays that were run in our laboratory prior to this study (see this thesis, Chapter 2), and were selected using the method as suggested by Lee and co-workers [152]. No elevated expression levels are expected for these genes.

Gene name	Description	Primers (5'- to 3'-)
An08g06940	histone, strong similarity to *A. nidulans* histone H4.1 [70]	Fw-ATCTTGCGTGACAACATCCA Rev-CACCCTCAAGGAAGGTCTTG
An08g05910	strong similarity to *A. nidulans* SagA which is involved in endocytosis and DNA repair [129]	Fw-CCAGGATGAAGAGTGGGAGA Rev-GCAGCTGGAGTGCTTCTTTC
An02g04120	similarity to *S. cerevisiae* Atx2 which is a Golgi transporter involved in manganese homeostasis [154]	Fw-TTTTCAGTCTGGCTGCTCCT Rev-CTGTTTTCCTGCATCGTGTG
An14g05050	strong similarity to *S. pombe* Dma1 which is a component of the spindle assembly point involved in mitotic division [183]	Fw-ACTCCAGAGGACAAGCAGGA Rev-GCAGACGCATGCTCTCAATA
An15g01860	malate synthase	Fw-TGATTAAGACGTGTCACCGC Rev-GGAGTGGGCATGTAGGTGTT
An01g00780	endo-xylanase B (*xlnB*)	Fw-CAACTTTGTCGGTGGAAAGG Rev-GGGTAGCCGTGTAGATATCG
An01g09960	β-xylosidase (*xlnD*)	Fw-TAATCTACGCCGGTGGTATC Rev-TTCTTGAGCGAAGAGGAATC
kana	kanamycin synthetase-encoding poly (A)-tailed synthetic gene	Fw-AGCATTACGCTGACTTGACG Rev-AGGTGGACCAGTTGGTGATT

Table 5 - Primers used in this study. The first four genes of this table are endogenous reference genes. 'Fw' and 'Rev' indicate forward and reverse primers.

Expression levels for malate synthase, whose expression is not influenced by D-xylose addition, also was measured. Finally, the transcriptional response upon the addition of D-xylose was measured by determining the transcript levels of the two genes, *xlnB* and *xlnD*. Primers were designed using the Primer3 web interface [218] and are given in Table 1. Primer design criteria were: length 20 to 22 base pairs; melting temperature 60 ± 1°C; GC-content 50 ± 5%. Amplicons ranged from 125 to 150 bp and had a melting temperature of 80 ± 5 °C.

Isolated total RNA was diluted in two steps to a concentration of 200 ng/µL. 1.00 µg of total RNA was spiked with 0.1 ng of the synthetic RNA transcript followed by DNaseI-treatment. cDNA was synthesized using the Omniscript reverse transcriptase kit (Qiagen). Four units of reverse transcriptase enzyme were added to the DNAse-treated total RNA in a final volume of 20 µL. cDNA synthesis reaction was at 37 °C and after one hour the synthesized cDNA was diluted 20-fold and stored at -20 °C until use. For qPCR measurements, PCR primers (final concentration 1.4 µM) and 4 µL of diluted cDNA were pipetted to 2× SYBR PCR mastermix (ABgene) using a CAS-1200 pipetting robot (Corbett Life Science).

A Rotorgene 3000 qPCR machine (Corbett Life Science) was used for thermocycling under the following conditions: 15 minutes at 95 °C; followed by 40 amplification cycles (denaturation, 15s at 95 °C; annealing, 15s at 58 °C; elongation, 20s at 72 °C). For each single run, non-template control samples for every primer pair used in that run were included. After the amplification cycles a melting step was performed. Quality control was done by inspection of the melting curve and samples with significant primer-dimer formation were not considered for analysis. The cycle threshold value and amplification efficiency were determined by the Rotorgene software using the 'comparative quantification' method (the cycle threshold value corresponds to the Rotorgene software 'take-off' value). The relative expression ratio of gene expression was calculated following Pfaffl [198]: ratio = $(E_{gene})^{\Delta Ct\ (pre\ -\ post)}$ / $(E_{kana})^{\Delta Ct\ (control\ -\ sample)}$. In this formula, E_{kana} denote the amplification efficiency of the synthetic kanamycin transcript and E_{gene} that of the gene for which the ratio is determined. ΔCt denotes the cycle threshold difference between pre- and post induced fermentor samples for the gene and *kana* transcripts.

Microarray processing

cDNA and cRNA synthesis and labeling and array hybridization were performed following the Affymetrix users' manual [3] using the One-cycle Target Labeling and Control Reagents Kit, and starting with 5 µg of total RNA as template material. Fifteen µg of fragmented and labeled cRNA was hybridized to custom-made *Aspergillus niger* arrays at 45 °C for 16 hours. Washing and staining was done using the Hybridization, Wash and Stain Kit (Affymetrix), using a GeneChip FS-450 Fluidics station and a Agilent G2500A GeneArray scanner. Scanned images were converted into .CEL files using MicroArray Suite software version 5 (Affymetrix).

Data analysis

For the 1,920 qPCR measurements obtained from the 5-week fermentor experiment, variance components were calculated by restricted maximum likelihood variance components analysis (REML sparse algorithm with Average Information optimization) [195] using GenStat 9.2 software (VSN International). Per gene, 3 REML analyses were run, using each gene's 'cycle threshold', 'amplification efficiency', and 'expression ratio' values as response variate. The random model, $y_{wfbrdqs} = \mu + \varepsilon_{week} + \varepsilon_{w.fermentor} + \varepsilon_{w.f.biomass} + \varepsilon_{w.f.b.RNA} + \varepsilon_{w.f.b.r.cDNA} + \varepsilon_{w.f.b.r.c.qPCR}$, was applied using $\varepsilon_{w.f.b.r.c.qPCR}$ as residual term.

For microarray data analysis, .CEL files of the individual array images were imported into GeneSpring 7.3 (Agilent technologies) using its RMA preprocessor to obtain RMA-normalized signal values for all arrays [124]. Raw and RMA-normalized array data are deposited in NCBI's GEO database [69], under series entry **GSE11405**. Probe sets with a RMA normalized signal below 37.7 - three times the lowest value detected over all arrays

- on all arrays were discarded, leaving 9,320 probe sets (64%). In comparison, when using the Affymetrix MAS5.1-software derived flag calls, on average 5,948 probe sets (41%) are called 'present' per array. For the 6 microarrays used in this study, statistically significant differentially expressed genes were determined using the limma package [239]. A Student's two-sample t-test between sorbitol and D-xylose arrays was executed, using empirical Bayesian statistics to moderate within-gene standard errors, Benjamini and Hochberg's 'false discovery rate' to correct for multiple testing [21], and an adjusted p-value < 0.05 to discriminate differentially expressed genes. To check for the influence of unequal sample size bias, testing was re-calculated with the inclusion of 4 additional sorbitol-grown samples derived from cultures grown identically (unpublished data), giving similar results.

Acknowledgements

We would like to thank J. Thissen of the Biometris expertgroup of Wageningen UR for performing the variance components analysis, and DSM Food Specialties for providing aid in fermentor optimization and use of the DSM *A. niger* microarrays.

Chapter 4

Transcriptional response of bioreactor-grown *Aspergillus niger* cultures towards three oils

Douwe van der Veen,
Willy A.M. van den Berg, and Leo H. de Graaff

Abstract

The industrially important fungus *Aspergillus niger* feeds naturally on decomposing plant material, for which it is equipped with a range of enzyme systems. Well-studied are cell wall-component degrading enzymes of the xylanolytic and pectinolytic systems. A significant proportion of plant material are lipids that may be used either as carbon or energy storage or as cellular building blocks. With 63 potential lipase-encoding genes in its genome, *A. niger* has the tools to degrade these extracellular lipids. In contrast to polysaccharide-degrading enzyme networks, not much is known about the signalling and regulatory processes that control lipase expression and activity in fungi both under laboratory and natural conditions.

A pulse of 1 mM of various oils was applied to four bioreactor-grown *A. niger* cultures to examine (i) whether *A. niger* responds at the level of gene transcription, (ii) at what time point this effect is detected most accurately, and (iii) whether differences between the response towards oils are observed. The triglyceride olive oil induces genes encoding peroxins and enzymes of fatty acid metabolism. A complex oil mixture extracted from wheat gluten, which is enriched for digalactosyl-diglycerides, induces genes encoding peroxins as well as enzymes of fatty acid metabolism, but with a different expression profile as compared to olive oil. Pure digalactosyldiglyceride, a proxy for plant membrane lipids, does not trigger a response of fatty acid metabolism-related genes or peroxins.

A pulse of 1 mM of lipids to a controlled bioreactor-grown *A. niger* culture is sufficient to induce a specific transcriptional response. Although olive oil properly can be used as 'default substrate' to study metabolic response upon oil induction, the natural response is more likely reflected in the oil mixture extracted from wheat gluten.

Background

The natural habitat of *Aspergillus niger* is soil, where this fungus is able to grow on decomposing plant material. It possesses a multitude of specialized enzymes that are secreted into its surroundings and which activities convert complex biomolecules into consumable substrates. Prime examples of such enzymes include pectinases and (hemi-) cellulases, enzymes that are also frequently used in the food and feed industry.

Plant material is mainly composed of carbohydrates and proteins, which for example for *Arabidopsis* leaves correspond to around 60% and 20% of dry weight, respectively [39]. Lipids make up 6% of the leave dry weight. However, the almost double energy content stored in lipids as compared to carbohydrates or proteins (37 kJ *vs.* 17 kJ per gram [244]) makes this rich source of energy not likely to be left unused. Different types of lipids are distinguished. Polar lipids form the matrix of plasma membranes that surround each cell and thus separate their internal from the external environment. Furthermore, plants may store energy for later use as neutral triacylglycerols or sterol esters in lipid bodies. Seeds, in plants the most common site for the accumulation of such lipid bodies, can contain oils as percentage of total seed weight ranging from 65% for walnuts to 2% for a non-oilseed species as peas [184].

The capacity of *A. niger* to utilize lipids as carbon source is evidenced by the presence of 63 putative lipase- or phospholipase- encoding genes in its genome [196]. Of these, 27 have a secretion signal. Lipolytic enzyme activity yields free fatty acids. These are taken up by the cell and are catabolized in a series of oxidation cycles at the β carbon atom, hence named 'β-oxidation pathway'.

Fatty acid metabolism has been studied in the closely related fungus *Aspergillus nidulans*, and regulators of fatty acid metabolism-related proteins have been identified although their precise mechanism of action is not fully understood [120, 258]. Much less is known about the regulation of lipases. *A. niger* produces lipases in the absence of lipids [234], but enzyme production is enhanced by addition of their substrate [131]. The type of lipids for optimal induction varies between strains [203]. Lipase activity is decreased in medium containing both glucose and lipids [131], which suggests a carbon-catabolite repression mediated control mechanism. Distinct lipolytic enzyme profiles are reported for 13 lipases and 5 phospholipases of *A. niger* [5, 6]. These studies demonstrate a difference in substrate specificity between the enzymes tested. For example, the product of gene An18g01090 solely has lysophospholipase activity, whereas gene An16g01880 encodes an enzyme that has phospholipase, lysophospholipase, and galactolipase activity [6].

For the yeast *Candida rugosa*, transcription profiles for 5 extracellular lipases differ for various culture conditions [111]. Likewise, 10 lipase genes of *C. albicans* show distinct transcription profiles in both lipid-containing and lipid-lacking media [112]. It is concluded that the different substrate specificities and distinct transcriptional profiles hint to the existence of distinct sets of lipases that might well be expressed to utilize different substrates.

Thus far, studies on the transcriptional response of eukaryotic micro-organisms towards fatty acid-related substrates are based on shake-flask grown cultures using one of two strategies. The first strategy is direct inoculation in medium containing the carbon source of interest (for example, see [228, 260]). A second strategy is to accumulate biomass on a neutral carbon source such as glycerol followed by transfer of the cells to fresh medium containing the carbon source of interest (for example, see [237]). Shake-flask cultivation is not optimal: important cultivation factors such as pH and dissolved oxygen concentration change uncontrollably over time and thus increase the experimental variation between replicate cultures. Combinatorial effects of growth parameters like medium composition or temperature might identify different sets of

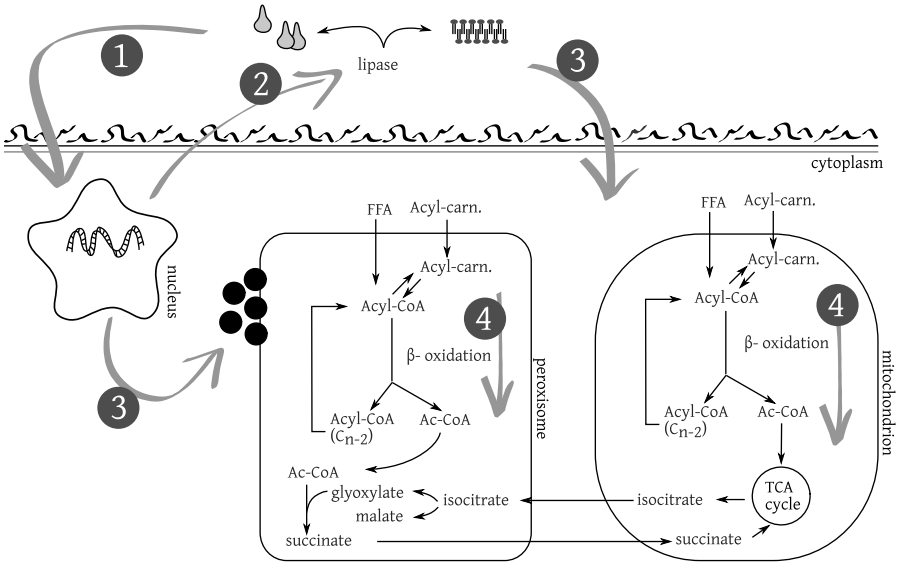


Figure 1 - Graphical representation of the *A. niger* cell, with the cytoplasm being separated from the outside environment by a membrane and cell wall. Abbreviations: FFA, free fatty acid, Carn., carnitine, TCA cycle, tricarboxylic acid cycle. CoA, Coenzyme A. Numbers indicate the four phases as described in the text: (1) substrate sensing, (2) extracellular lipid degradation, (3) uptake of fatty acids and organelle adaptation, and (4) fatty acid metabolism.

differentially expressed genes in repeated experiments [144]. Furthermore, gene expression levels are affected by the organisms growth rate, which makes comparison of directly inoculated cultures difficult [44, 144].

Many of the disadvantages of shake-flask cultivation are bypassed in bioreactors, where cultivation conditions can be both monitored and controlled. The aim of this study was to determine whether our bioreactor-based experimental set-up facilitates the study for global transcriptional responses towards induction by lipids. For this purpose the global transcript response for distinct lipid mixtures was determined at three time points after induction. The transcriptional response was analysed by a qualitative data analysis. This data analysis was guided by four phases describing physiological processes of fatty acid metabolism: (i) sensing of the substrate, and response to it, (ii) degradation of the substrate, (iii) transport into the cell or organelle, and organelle adaptation, and (iv) metabolic processing of the substrate (Figure 1).

Results and discussion

Lipids

Refined olive oil has been used in this study as reference inducer. Olive oil is a mixture of triacylglycerols with oleic acid (C18:1) and linoleic acid (C18:2) as primary fatty acids (the molecular structure of the oils used in this study are shown in Figure 2). The response towards plant membrane lipids was studied by induction with purified galactolipids. In membranes of photosynthetic tissues, phospholipids have been partly replaced by galactolipids to reduce dependence on the limited availability of inorganic phosphate in many soil types [65]. Around 75% of lipids are galactolipids in photosynthetic tissues, whereas in non-photosynthetic tissues this percentage is around 20% [66]. Galactolipids are diacylglycerols with two galactose moieties bound, and the

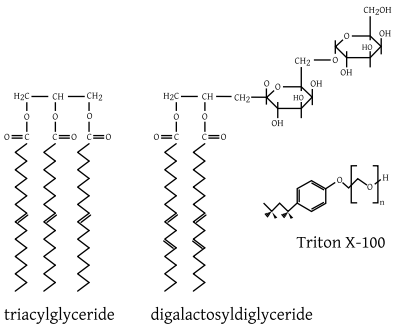


Figure 2 - Schematic representation of oil molecules used. Left: 1,2,3-tri (*cis*-9-octadecenoyl) glycerol, or oleic acid triglyceride. Middle: 1,2-dilinoyl-3-O-(α-D-galactopyranosyl-(1>6)-O-β-D-galactopyranosyl)-sn-glycerol, or digalactosyldiglyceride. Right: 4-octylphenol polyethoxylate. For Triton X-100, n is 9 or 10.

majority of fatty acids in wheat flour has two unsaturated bonds (C18:2) [14]. A third oil mixture aimed to mimic the natural response of *A. niger* as it encounters lipid-containing plant material. This oil mixture was extracted from wheat kernels. The emulsifying agent used in this study is the non-ionic surfactant Triton X-100.

Phase 1: Substrate sensing and cellular response

The signalling mechanism that evokes a response to the presence of lipids in the environment of *A. niger* is unknown.

In *Saccharomyces cerevisiae*, peroxisome-related proteins and fatty acid metabolic enzymes are induced by fatty acids of varying chain lengths [236], with the degree of saturation of the fatty acid chain not being of influence [93]. The combined action of two $Zn(II)_2Cys_6$-class transcriptional activators, Pip2p and Oaf1p, mediates transcription of genes encoding peroxisomal proteins and enzymes of the fatty acid β-oxidation [132]. These proteins bind as a heterodimer to oleate response elements found in the promoter region of genes under their control [215] and act as exclusive sensor for a variety of fatty acids via the responsive partner, Oaf1p [19, 216]. Pip2p-Oaf1p in turn is directly governed by Adr1p, a transcriptional activator of genes controlling aerobic metabolism of non-glucose carbon sources such as ethanol, glycerol, or fatty acids [217]. The Adr1p regulon is subject to glucose repression [293] and is only active when no glucose is present.

Orthologs of the Pip2p-Oaf1p complex are present in the genomes of yeasts but not in the genomes of other eukaryotes including filamentous fungi. This suggests that for these organisms a different regulatory mechanism is present. In the last decade, four factors involved in transcription of fatty acid related genes have been identified in *A. nidulans* by Hynes and co-workers [120].

The *facB* gene encodes a transcription factor that is induced by acetate [134, 258]. This transcription activator controls among others the genes encoding acetyl-CoA synthetase (*facA*), cytosolic carnitine acyltransferase (*acuC*), and the glyoxylate shunt enzymes isocitrate lyase (*acuD*) and malate synthase (*acuE*) [256, 257]. Disruption of *facB* blocks growth on acetate as sole carbon source but maintains its ability to grow on fatty acids as sole carbon source. The *A. niger facB* ortolog, An08g06580, is induced up to 1.4-fold in time in all four fermentations.

Recently, two $Zn(II)_2Cys_6$-type transcription factors were identified in *A. nidulans*: FarA and FarB [120]. Both transcription factors bind to the same DNA sequence, which core of 5'-CCGAGG-3' is found in the upstream region of 40 of 49 genes encoding proteins involved in fatty acid metabolism [120]. All but one gene that are under FacB control contain this conserved sequence (the exception is the cytoplasmic carnitine

acetyltransferase *facC*, which is under FacB control [242] but does not have this sequence in its upstream region). Deletion of *farA* eliminates growth on both short- and long-chain fatty acids, while a mutation in *farB* eliminates growth on short-chain fatty acids but not on long-chain fatty acids. In *A. niger*, the *farA* ortholog is up to 4-fold induced on wheat oil and 1.6-fold on olive oil, while transcript levels remain below 1.4-fold on either DGDG and in the negative control fermentation. Its expression level reaches a maximum at 30 minutes after induction and declines to pre-induced levels after 2 hours. The *farB* expression level is slightly elevated - up to 1.3-fold compared to non-induced levels - in all fermentations but the negative control fermentation. This elevated expression occurs at 1 hour after induction only.

Based on the observation that, in a *farA* deletion strain, transcript levels of a reporter gene under control of the carnitine acetyltransferase promoter are elevated under non-induced conditions, a repressing function for FarA in the absence of inducer was suggested [120]. The transcription profile for the *farA* ortholog of *A. niger* in our experiment suggests transcription of *farA* prior to transcription of *farB*, which is difficult to explain when FarA would act as repressor. Although the expression levels of the *farB* ortholog are low, the consistent elevated expression at 1 hour might hint to a recruitment model where FarA binding is required prior to binding of the FarB transcription factor.

A third gene encoding a protein with regulatory function as described in [120], *scfA*, had no altered transcript levels under the conditions examined.

In summary, only the FarA transcription factor is clearly induced by olive oil and wheat oil. This induction reaches a maximum already at 30 minutes after induction. FarB expression levels are low but an apparent maximum expression level is reached at 1 hour after induction (1.3-fold). Both *farA* and *farB* do not have the consensus binding sequence to which they bind in their upstream region [120]. The observed expression pattern suggests that *farA* expression is elevated prior to that of *farB*, but it remains unclear whether FarA directly affects *farB* transcription.

Phase 2: extracellular lipid degradation

The *A. niger* genome contains 63 genes annotated as lipase-encoding. For 20 of these genes, raw signal intensities were found below the threshold setting and were not further analysed. Table 1 presents the expression values for the remaining genes relative to the pre-induced sample of each fermentation. The majority of genes have no elevated expression; however, 6 lipase-encoding genes are over 2-fold differentially expressed in at least one fermentation.

			DGDG				Olive oil				Non-induced				Wheat oil			
Gene name	**Activity**	**Secr.**	**0**	**0.5**	**1**	**2**	**0**	**0.5**	**1**	**2**	**0**	**0.5**	**1**	**2**	**0**	**0.5**	**1**	**2**
An01g05620			1.0	1.1	1.3	1.2	1.0	1.5	1.2	1.3	1.0	1.1	1.1	1.1	1.0	1.3	1.2	1.1
An01g07280			1.0	1.0	0.9	0.9	1.0	1.1	0.8	0.9	1.0	1.2	1.2	1.2	1.0	0.8	0.9	0.8
An02g04040			1.0	0.8	0.9	0.9	1.0	1.1	1.1	0.9	1.0	1.2	1.0	0.9	1.0	1.0	1.0	1.0
An02g04680			1.0	0.9	1.1	1.5	1.0	0.9	1.1	1.5	1.0	1.1	1.2	1.2	1.0	1.2	1.8	**2.1**
An02g09690		yes	1.0	1.0	1.0	1.2	1.0	1.0	1.2	1.1	1.0	1.2	1.1	1.0	1.0	1.1	1.2	1.1
An02g11290			1.0	1.1	1.3	1.2	1.0	1.1	1.1	1.2	1.0	1.0	1.4	1.2	1.0	1.3	1.4	1.3
An02g13220	P Y G[§]	yes	1.0	1.3	1.3	1.9	1.0	**2.2**	1.9	**2.3**	1.0	**2.0**	1.8	1.9	1.0	1.5	1.3	1.6
An02g13440		yes	1.0	0.8	1.0	0.7	1.0	1.0	0.9	0.7	1.0	1.2	0.9	0.8	1.0	1.1	1.4	1.0
An03g02100			1.0	1.0	1.0	1.1	1.0	0.9	0.9	1.1	1.0	0.9	1.1	1.0	1.0	1.0	1.2	1.2
An03g02820			1.0	1.0	1.0	1.0	1.0	0.9	1.0	1.0	1.0	1.1	1.0	1.2	1.0	1.4	1.4	1.2
An03g06560	L P Y G[‡]	yes	1.0	0.9	1.0	0.9	1.0	1.1	1.0	1.0	1.0	1.1	1.0	1.0	1.0	1.1	1.1	1.2
An04g00510			1.0	1.5	1.3	1.2	1.0	1.2	1.0	1.1	1.0	1.1	1.3	1.1	1.0	1.2	1.2	1.0
An04g03360			1.0	0.7	0.7	0.7	1.0	0.6	0.7	0.6	1.0	0.9	0.8	0.6	1.0	0.6	1.1	0.7
An07g00440	L P Y G[‡]	yes	1.0	1.1	1.1	1.2	1.0	**2.4**	1.6	1.9	1.0	**2.7**	1.4	1.1	1.0	**2.1**	**2.1**	**2.2**
An07g02240			1.0	1.0	1.0	0.9	1.0	0.9	1.0	0.9	1.0	1.3	1.0	1.0	1.0	0.8	1.0	1.0
An07g04200	P[‡]	yes	1.0	1.1	0.9	0.9	1.0	0.9	0.9	0.9	1.0	0.9	0.8	0.9	1.0	0.8	1.0	1.0
An08g06320			1.0	1.1	1.2	1.1	1.0	1.3	1.3	1.8	1.0	0.8	1.0	1.3	1.0	0.8	0.9	1.3
An08g08490			1.0	0.8	1.0	1.1	1.0	0.7	0.8	1.0	1.0	0.8	1.0	1.1	1.0	0.5	1.0	1.4
An09g01240		yes	1.0	0.7	0.8	0.8	1.0	1.4	0.7	0.8	1.0	2.0	1.1	0.8	1.0	1.6	0.7	0.8
An09g02180		yes	1.0	1.4	1.2	1.5	1.0	1.3	1.5	1.7	1.0	0.9	1.1	1.1	1.0	1.1	**2.8**	**2.8**
An09g02270	P Y [‡]	yes	1.0	1.0	1.0	1.0	1.0	1.0	1.2	1.2	1.0	1.0	1.0	1.0	1.0	0.9	1.2	1.3
An09g06060		yes	1.0	0.9	0.9	0.8	1.0	1.0	0.9	1.0	1.0	0.9	0.9	0.9	1.0	0.8	0.8	0.8
An09g06390			1.0	1.1	1.1	1.0	1.0	1.0	1.0	0.9	1.0	0.9	1.1	1.1	1.0	0.8	1.2	0.9
An11g00100			1.0	1.2	1.1	0.7	1.0	1.8	1.2	0.9	1.0	1.0	1.0	0.7	1.0	**3.0**	1.9	1.1
An11g00440			1.0	1.1	1.1	1.2	1.0	1.0	1.0	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.1
An11g01640			1.0	1.2	1.0	0.9	1.0	1.0	0.9	1.0	1.0	1.0	1.1	0.9	1.0	0.9	0.7	0.7
An11g03380			1.0	1.1	1.1	1.1	1.0	1.1	1.2	1.5	1.0	0.9	1.1	1.1	1.0	1.8	1.9	1.7
An11g11010			1.0	1.0	1.0	1.0	1.0	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.8
An12g00630			1.0	1.0	1.2	0.9	1.0	1.0	1.0	1.0	1.0	0.9	1.0	1.1	1.0	0.8	1.1	1.3
An12g03120			1.0	1.2	1.0	1.1	1.0	1.2	1.2	1.1	1.0	1.2	1.3	1.1	1.0	1.0	1.0	1.1
An12g06560	L P Y [‡]	yes	1.0	1.0	0.9	1.0	1.0	1.1	1.1	1.0	1.0	1.1	1.1	1.0	1.0	1.0	1.1	1.1
An12g06690			1.0	1.0	1.0	1.0	1.0	1.0	0.9	1.1	1.0	1.1	1.0	1.0	1.0	1.1	1.0	1.0
An12g07760			1.0	1.1	1.1	1.0	1.0	1.1	1.1	1.0	1.0	1.1	1.1	1.0	1.0	1.1	1.2	1.0
An13g03600			1.0	1.2	1.0	1.3	1.0	0.9	1.1	1.1	1.0	1.1	1.1	1.0	1.0	0.9	1.1	1.1
An14g00860	L P Y [‡]	yes	1.0	1.1	1.1	1.2	1.0	0.9	0.9	1.2	1.0	0.9	1.0	1.1	1.0	0.9	1.0	1.4
An15g03950			1.0	1.0	0.8	0.9	1.0	0.8	0.8	0.7	1.0	0.8	1.0	1.0	1.0	0.7	0.7	0.7
An15g07040			1.0	0.8	1.1	1.2	1.0	1.2	1.1	1.2	1.0	1.4	1.4	1.4	1.0	0.6	1.1	1.1
An16g01880	P Y G[§]	yes	1.0	0.9	**3.5**	**9.9**	1.0	1.1	**3.7**	**13.8**	1.0	1.2	1.4	**3.3**	1.0	1.1	**10.4**	**30.8**
An16g03700		yes	1.0	1.0	1.6	1.1	1.0	0.8	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.8	1.7	1.0
An16g08870	L[‡]	yes	1.0	0.9	0.9	1.0	1.0	1.1	1.0	1.1	1.0	1.1	1.1	1.1	1.0	1.1	1.2	1.1
An18g01090	Y[§]	yes	1.0	0.8	0.8	0.8	1.0	0.9	0.7	0.7	1.0	1.3	1.0	0.9	1.0	0.7	0.7	0.8
An18g06580		yes	1.0	1.0	1.0	0.9	1.0	1.0	1.0	0.9	1.0	1.0	1.0	1.0	1.0	0.9	1.0	1.1

Table 1 - The 43 lipases identified in the genome of *A. niger* CBS513.88 with raw expression signal value over 50 for at least 1 microarray. Transcript levels are relative to the signal of the non-induced sample taken from that fermentor (T=0). Bold font indicate over 2-fold expression, bold underlined font indicates over 3-fold expression. Enzyme activity: [§] as described in [6]; [‡] as described in [5]. *Secr.* indicates the presence of a secretion signal.

Of these, the gene with highest expression level is An16g01880, which is over 10-fold over-expressed in DGDG, olive oil, and wheat oil-induced samples in a time-dependent manner. Its encoded enzyme has phospholipase A, lysophospholipase, and galactolipase activity [6]. Expression of this gene is also elevated for the Triton control fermentation (but only up to 2.8-fold). However, around 3-fold increased expression is also observed for this gene one hour after induction with the sugar D-xylose (unpublished results). The observation that transcription of this gene not only responds to the oils used in this study but also to conditions not expected to induce expression of this lipase-encoding gene, suggests that its expression levels can as well be the result of time-dependent expression.

Summarizing Table 1, in our experiments only a limited response of lipase-encoding genes is found in *A. niger* upon induction with oil, and no major differences are observed between the four induction conditions.

Phase 3: uptake of fatty acids

The mechanism that facilitates the uptake of fatty acids into eukaryotic cells is not fully understood, but experimental data suggests that two distinct processes might play a role: passive diffusion through the lipid bilayer membrane [296] and facilitated transport via a membrane-embedded protein or protein complex [94, 241]. In *Escherichia coli*, transport of long-chain fatty acids across its cell envelope requires the FadL and FadD proteins (reviewed in [63]). FadL, located in the outer membrane, is a transport protein that channels its fatty acid ligands to the periplasmic space. This periplasmic space is slightly acidic; the long-chain fatty acids become protonated, diffuse into the inner membrane, and flip to the cytosolic side of this membrane by passive diffusion. The FadD protein is an acyl-CoA synthetase that is located in the cytosol. Upon binding of ATP, this protein moves to the inner membrane and activates fatty acids by linking them to Coenzyme A (CoA). After activation, these fatty acids are further metabolized for fatty acid biosynthesis or degradation.

In higher eukaryotic systems other classes of fatty acid transport proteins have been described: fatty acid translocase (FAT) [96]; plasma membrane-bound fatty acid-binding protein (FABPpm) [243]; and adipocyte fatty acid transport protein (FATP) [224]. However, in contrast to the bacterial transporter FadL, all of these proteins have additional functions next to transport of fatty acids [63]. For example, FATP proteins also possess fatty acid synthetase activity.

Three proteins appear to be principal players in protein-mediated fatty acid import and activation in *S. cerevisiae*. Fat1 encodes a membrane-bound long-chain fatty acid transport protein that is 33% identical and 54% similar to adipocyte FATP [75].

Disruption of this gene impairs growth on fatty acids as carbon source and decreases uptake of a fluorescent fatty acid analogue. However, as Fat1p is localized in both endoplasmic reticulum and peroxisomal matrix rather than the plasma membrane [283], the debate regarding the primary function of this protein in fatty acid uptake is ongoing. Two other long-chain fatty acid synthetases, encoded by Faa1 and Faa4, also play a role in fatty acid import [76]. A double deletion of these genes results in 99% reduction of uptake of 14:0 and 16:0 fatty acids [272]. This reduction is explained as a failure to convert fatty acids into acyl-CoA esters, which disrupts the concentration gradient of fatty acids over the membrane. In this view, the import of fatty acids is closely linked to fatty acid activation like for *E. coli* [76, 107]. Next to Faa1 and Faa4, *S. cerevisiae* encodes 2 more acyl-CoA synthetases but their cellular roles are unclear. Disruption of the Faa2 and Faa3 genes does not affect the growth on exogenously supplied fatty acids, which suggests that the encoded enzymes access only fatty acids synthesized within the cell [126].

Once the activated fatty acids are inside the cell, acyl-CoA binding proteins (ACBPs) are thought to regulate the size and composition of the cytosolic acyl-CoA pool [226]. These activated fatty acids must be transported to the peroxisome or mitochondria for further processing. To enter these organelles, the activated fatty acids are transported over the peroxisomal or mitochondrial membrane with the aid of carnitine acyltransferases. Alternatively, fatty acids can be imported into peroxisomes without the involvement of carnitine. In *S. cerevisiae*, this process is facilitated by an ABC transporter complex composed of Pxa1 and Pxa2 [103].

An ortholog of the yeast Fat1p fatty acid transport protein is An07g09190. Expression of this gene is 2-fold increased in olive oil-induced mycelium compared to non-induced mycelium. Also, this gene is up to 5-fold induced on wheat oil 30 minutes after induction. A second gene with similarity to Fat1p, An02g03330, has no elevated transcript levels in any fermentation.

No orthologs of the yeast Faa3p and Faa4 genes are found in *A. niger*, but orthologs of Faa1p and Faa3p are present in its genome. An16g05150, the Faa1p ortholog, has no elevated gene transcript levels in any of the samples. However, the Faa2p ortholog, An09g06740, has 3-fold increased transcript levels upon olive oil induction and 8-fold increased levels upon wheat oil induction. The putative acyl-CoA binding protein in *A. niger*, encoded by An03g01800, is not induced in any of the examined conditions.

Two carnitine acyltransferases are annotated in the *A. niger* genome: An18g01590 and An03g03360. This latter gene encodes the ortholog of the *A. nidulans* carnitine acyltransferase, AcuH [158]. Both genes have elevated expression of 5- fold and 8-fold on

wheat oil, respectively. When induced with olive oil, transcription for these genes is elevated as well (2.8-fold). No elevated transcript levels are observed in either the DGDG-induced mycelial samples or in the negative control.

A further two proteins are orthologous to the *S. cerevisiae* Agp2p plasma membrane carnitine transporter: An04g00530 and An04g09620. An04g00530 expression levels at 30 minutes after induction are 2.5-fold increased on wheat oil and 1.8-fold on DGDG and olive oil. One hour after induction, expression has returned to pre-induction levels. No elevated expression is observed in the non-induced fermentation. The transcription of An04g09620 does not alter in any of the conditions tested.

The *A. niger* orthologs of the Pxa1p/Pxa2p ABC transporter complex of *S. cerevisiae*, An01g03680 and An08g05780, have elevated transcript levels of 2- to 3-fold on olive oil and wheat oil, respectively.

Phase 3 - organelle adaptation

Microbodies are a family of single-membrane-surrounded organelles of which glyoxysomes, peroxisomes and glycosomes are members. These organelles compartmentalize various oxidative reactions and thus shield the cell from the damaging effect of reactive oxygen species formed such as hydrogen peroxide. The various types of microbodies are distinguished on the basis of their enzymatic content and marker enzymes. Enzymes imported into microbodies share similar trafficking signals: the peroxisomal targeting signals PTS1 and PTS2 [245]. Furthermore, microbody family members are able to degrade fatty acids [226].

Peroxisomes are essential in mammals and plants [136, 279] but appear to be dispensable in fungi under certain conditions. For instance, peroxisome-deficient *S. cerevisiae* mutants are unable to grow on fatty acids as sole carbon source but are perfectly viable when maintained on glucose [73]. Notwithstanding these differences, the mechanisms of peroxisome assembly and biosynthesis have been evolutionary well conserved [151]. This conservation has allowed for a unified and species-independent nomenclature for peroxisome biosynthesis and assembly genes: *pex* genes or peroxins [64]. A systematic analysis identified the currently known peroxin genes for of 17 yeast and fungal species, but not for *A. niger* [136]. The in this study identified peroxin-encoding genes were used to identify orthologs in *A. niger*. No orthologs were found for *pex4*, *pex14/17*, *pex11B*, and *pex24*, although these genes are present in *A. nidulans*, *A. fumigatus*, and *Penicillium chrysogenum*. The twenty identified peroxins were examined in this study.

		DGDG				Olive oil				Non-induced				Wheat oil			
Gene name	**Peroxin**	**0**	**0.5**	**1**	**2**	**0**	**0.5**	**1**	**2**	**0**	**0.5**	**1**	**2**	**0**	**0.5**	**1**	**2**
An16g04880	Pex1	1.0	1.2	1.1	1.1	0.8	1.0	1.1	1.1	1.0	1.0	1.0	1.0	1.0	**2.2**	1.5	1.1
An18g04150	Pex2	1.0	1.3	1.1	0.9	1.1	1.5	1.5	1.6	1.0	1.2	1.2	1.1	1.0	**3.5**	**2.1**	1.4
An17g01440	Pex3	1.0	1.2	1.1	0.9	0.9	1.1	1.1	1.2	1.0	1.2	1.1	1.0	1.0	1.4	1.4	1.1
An16g07480	Pex5	1.0	1.0	1.1	1.1	1.1	1.1	1.3	1.2	1.0	1.1	1.0	1.1	1.0	1.4	1.7	1.2
An16g02010	Pex6	1.0	1.1	1.2	1.0	1.3	1.4	1.7	1.7	1.0	1.3	1.2	1.1	1.0	**3.4**	**2.1**	1.1
An01g13690	Pex7	1.0	1.1	1.2	0.9	1.0	1.1	1.3	1.6	1.0	1.0	1.1	1.0	1.0	**2.2**	**2.0**	1.4
An12g03650	Pex8	1.0	1.0	1.1	1.1	1.2	1.4	1.3	1.3	1.0	1.3	1.2	1.2	1.0	1.6	1.6	1.4
An13g00800	Pex10	1.0	1.3	1.0	1.0	0.9	1.3	1.4	1.4	1.0	1.1	1.0	1.0	1.0	**4.8**	**2.3**	1.7
An11g02590	Pex11	1.0	1.0	1.1	1.0	1.0	1.2	1.4	1.6	1.0	1.1	1.2	1.1	1.0	1.8	1.9	1.4
An04g09150	Pex11C	1.0	1.2	1.0	1.1	0.9	1.0	1.0	0.9	1.0	1.0	1.0	0.9	1.0	1.0	1.0	1.0
An15g06690	Pex12	1.0	0.9	1.0	1.0	0.9	1.2	0.9	1.0	1.0	0.9	1.0	0.9	1.0	1.5	0.7	0.9
An16g07610	Pex13	1.0	1.0	1.1	1.1	0.9	1.2	1.3	1.5	1.0	0.9	1.0	1.0	1.0	**2.0**	1.9	1.3
An16g04020	Pex14	1.0	1.2	1.2	1.1	0.9	1.2	1.1	1.1	1.0	1.1	1.0	1.0	1.0	1.6	1.6	1.0
An07g10240	Pex16	1.0	1.2	1.3	1.1	0.9	1.6	1.7	**2.1**	1.0	1.2	1.2	1.0	1.0	**3.1**	**2.4**	1.8
An02g06680	Pex19	1.0	0.8	1.2	1.2	1.1	0.8	1.1	1.0	1.0	1.1	1.0	1.1	1.0	1.2	1.1	1.2
An04g00370	Pex20	1.0	1.2	1.2	1.1	1.0	1.2	1.3	1.4	1.0	1.1	1.0	1.1	1.0	**2.0**	1.5	1.4
An01g10460	Pex22-like	1.0	1.0	0.9	1.0	1.0	1.0	0.9	1.0	1.0	1.0	1.0	0.9	1.0	1.3	1.2	1.1
An13g00680	Pex23	1.0	0.9	1.0	1.0	1.0	0.9	1.0	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
An16g03610	Pex23-like	1.0	1.0	0.9	0.9	0.9	1.0	1.0	1.1	1.0	0.9	0.9	1.0	1.0	1.0	1.0	1.1
An04g09060	Pex26	1.0	1.1	1.1	1.0	1.0	1.2	1.2	1.1	1.0	1.1	1.0	1.0	1.0	1.9	1.5	1.0

Table 2 - This table depicts relative expression data for the 20 identified peroxins. Transcript levels are relative to the signal of the non-induced sample taken from that fermentor (T=0). Bold font indicates over 2-fold expression, bold underlined font indicates over 3-fold expression.

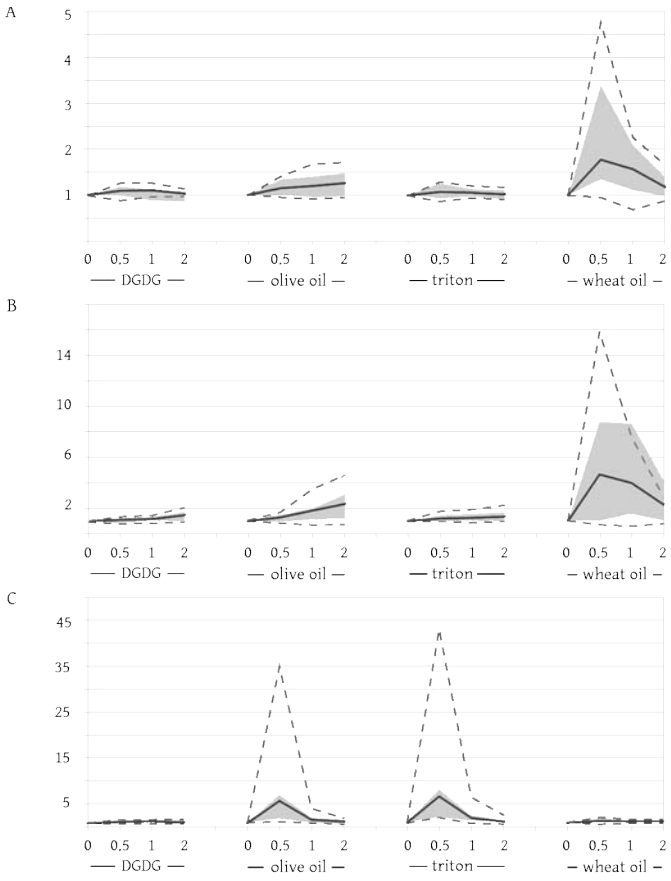


Figure 3 - Gene expression profiles of the (A) peroxins of Table 2, (B) genes encoding enzymes of the β-oxidation pathway of Table 3, and (C) genes that respond to Triton. Expression level ratio is relative to the non-induced level of each fermentation, which is set to 1. The thick black line indicates the median transcript level. The grey area encompasses the transcript level boundaries of the 75 % of genes closest to the median. The dotted lines bound the most extreme transcript level for each fermentation.

The relative expression levels for 16 peroxins involved in the biogenesis of the peroxisomal membrane and matrix protein import as well as for 4 peroxins involved in peroxisome proliferation are analyzed (Table 2, Figure 3). Genes encoding peroxins involved in peroxisome biogenesis and matrix protein import have elevated transcript levels on olive oil and wheat oil. In the DGDG and non-induced control fermentations, transcript levels remain unaltered compared to each non-induced time point sample.

The expression pattern for these genes differs between the oils: expression levels in the olive oil-induced samples raise over time, while in the wheat oil-induced culture a maximum expression level is reached after 30 minutes. This maximum then is followed by a decline to pre-induced levels after 2 hours.

Of the 4 peroxins involved in peroxisome proliferation, only *pex11* has transcription levels elevated up to 1.9-fold when induced with olive oil or wheat oil. Four peroxin-encoding genes have no elevated expression in any fermentor: 3 peroxins involved in peroxisome proliferation (*pex11C*, *pex23*, and *pex23-like*), and the ortholog of Pex19, An02g06680. This latter peroxin is involved in peroxisomal membrane formation, and interacts with Pex3 [229]. The *A. niger* Pex3 ortholog is induced on both oils, however.

Other genes that encode peroxisome-related proteins have elevated expression levels similar to those for the peroxins: levels are increased on olive oil and wheat oil. No elevated transcript levels are detected for the fermentation induced with pure DGDG and the non-induced fermentation. *S. cerevisiae* Ant1p is a carrier protein that transports ATP into the peroxisome [273]. Expression of its encoding ortholog, An01g04690, is up to 9-fold elevated on wheat oil and up to 2.3-fold on olive oil. A peroxisomal ABC transporter, encoded by An01g03680, is up to 3-fold induced both on olive oil and wheat oil. A third peroxisomal transporter, encoded by the ortholog of the *pmp47* gene of *C. Boidinii* [185], An09g06780, is 4-fold induced by wheat oil and 2.8-fold by olive oil. Finally, An08g10110 is induced both on olive oil (3-fold) and wheat oil (4-fold). This gene encodes the ortholog of a putative non-specific lipid transfer protein of *C. tropicalis*, a protein that is located in the peroxisomal matrix but of which the exact function is not determined [249].

The fungus responds to the addition of olive oil or wheat oil by an increased production of peroxin transcripts. Also, peroxisome transporter levels are increased in these cultures. These data suggest that *A. niger* responds to these lipids by alteration of its physiology to accommodate the processing of fatty acids.

Phase 4: fatty acid metabolism

Degradation of CoA-activated acyl chains involves cycles of four enzyme-catalyzed steps, where every cycle releases one molecule of acetyl-CoA (Figure 4). Whereas β-oxidation in *S. cerevisiae* is restricted to peroxisomes [107, 148], in non-yeast fungi, β-oxidation pathways are located in both peroxisomes and mitochondria [166]. Genes involved in both peroxisomal and mitochondrial β-oxidation in *A. nidulans* have been identified in recent years [120, 166]. Disruption of *foxA*, which encodes a bifunctional peroxisomal enzyme with both enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase activity, totally disabled growth of this organism on the long-chain fatty acid, erucic acid (C22:1),

whereas growth on short-chain fatty acids was not affected [166]. The presence of a mitochondrial β-oxidation pathway was evidenced by disruption of the *echA* gene, which encodes a mitochondrial short-chain enoyl-CoA hydratase [166]. Next to EchA, the protein encoded by *scdA* (short-chain acyl-CoA dehydrogenase) [165] and the products of *hadA* and *mthA* (respectively encoding 3-hydroxyacyl-CoA dehydrogenase and acetyl-CoA C-acyltransferase) are proposed to take part in the mitochondrial β-oxidation pathway [121].

The first step in peroxisomal β-oxidation is catalyzed by acyl-CoA oxidases. These FAD-containing enzymes transfer electrons directly to oxygen to generate H_2O_2. In contrast, mitochondrial acyl-CoA dehydrogenases pass electrons via electron transferring flavoprotein (ETF) onto the respiratory chain [85, 202]. Acyl-CoA oxidases are dimeric proteins while acyl-CoA dehydrogenases are tetrameric, but both are members of the same superfamily [27]. Mammals have three acyl-CoA oxidases (ACOX1, -2, and -3) while *S. cerevisiae* has only one (Fox1p) [202]. The *A. nidulans* ortholog of Fox1p is AN6752, but no *A. niger* strain CBS513.88 ortholog has been detected in its genome sequence using BlastX or BlastN techniques. The genome of *A. niger* strain ATCC1015 (http://genome.jgi-psf.org/Aspni5/Aspni5.home.html) does contain an ortholog for this acyl-CoA oxidase (located on contig 5, 2262620 - 2264873). Seven acyl-CoA dehydrogenases are present in the *A. niger* genome (Table 3).

In the second step of β-oxidation, the C=C double bond between the α and β position of the acylgroup is hydrolyzed by 3-hydroxyacyl-CoA hydrolase activity. In *A. nidulans*, the above mentioned bifunctional enzyme (FoxA) catalyzes this reaction in peroxisomes while the same reaction in mitochondria is catalyzed by EchA.

3-Hydroxyacyl-CoA is oxidized in the third step of β-oxidation by a dehydrogenase to form 3-oxoacyl-CoA. In peroxisomes, the bifunctional FoxA enzyme also catalyzes this reaction. In mitochondria in *A. nidulans*, the *hadA* gene encodes the enzyme catalyzing this step; an ortholog in *A. niger* is present.

The fourth and final step of β-oxidation of fatty acids is the thiolysis of 3-oxoacyl-CoA. Acetyl-CoA is formed in this reaction, which will enter the TCA cycle or fulfil any of its other ubiquitous roles in the cell. The also formed acyl-CoA molecule, now shortened by two carbon atoms, will enter another round of β-oxidation. Acetyl-CoA C-acyltransferases catalyze this thiolysis, and 5 genes encoding putative homologous enzymes are present in the *A. niger* genome.

In total 19 genes are identified in the *A. niger* genome whose products are putative enzymes of the β-oxidation pathway (Table 3). The expression profiles for these genes

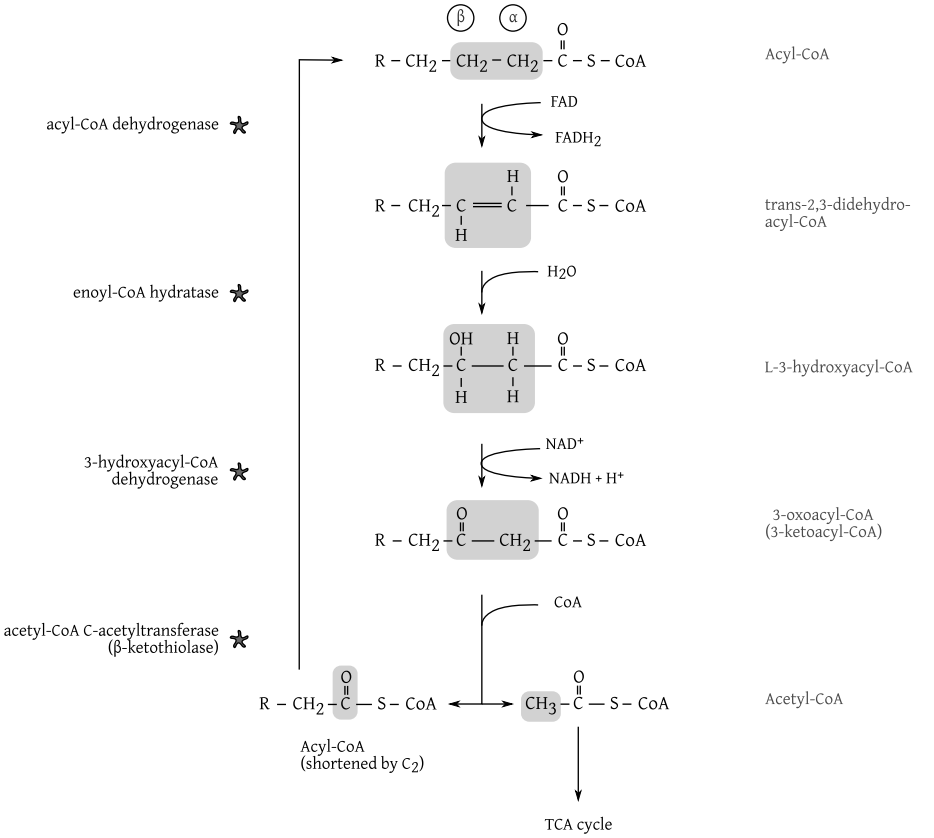


Figure 4 - The four steps in degradation of fatty acids by the β-oxidation pathway in mitochondria. In peroxisomes, the first step is catalyzed by acetyl-CoA oxidases, where electrons are transferred to molecular oxygen instead of ETF. Enzymes that catalyze each step are indicated by an asterisk.

are given in Figure 3. No elevated expression levels for any of these 19 genes of the β-oxidation pathway is observed for either the non-induced control fermentation or the digalactosyldiglyceride induced fermentation. However, induction by olive oil and/or wheat oil does result in differential expression.

Three of seven predicted acyl-CoA dehydrogenase encoding genes in the *A. niger* genome have over two-fold induced expressed on olive oil compared to non-induced levels. For gene An04g03290, a putative long-chain acyl-CoA dehydrogenase, expression levels are over two-fold increased already after 30 minutes after induction with olive oil. Wheat oil appears a more potent inducer of acyl-CoA dehydrogenase encoding genes: four genes

Gene name	Description	DGDG				Olive oil				Non-induced				Wheat oil			
		0	0.5	1	2	0	0.5	1	2	0	0.5	1	2	0	0.5	1	2
An11g00400	acyl-CoA dehydrogenase	1.0	1.0	1.1	1.0	1.0	1.5	2.0	**3.1**	1.0	1.1	1.1	1.1	1.0	**15.8**	**7.4**	**2.8**
An09g03730	acyl-CoA dehydrogenase	1.0	1.1	1.1	1.0	1.0	1.0	1.2	1.5	1.0	1.2	1.2	1.3	1.0	1.4	1.6	1.2
An17g01150	acyl-CoA dehydrogenase (medium-chain)	1.0	1.2	1.2	1.3	1.0	0.8	1.4	1.9	1.0	1.3	1.2	1.4	1.0	**4.0**	**5.4**	1.8
An04g03290	acyl-CoA dehydrogenase (long-chain)	1.0	1.3	1.4	1.3	1.0	**2.5**	**3.8**	**4.8**	1.0	1.4	1.5	1.5	1.0	**4.9**	**6.6**	**4.7**
An13g03940	acyl-CoA dehydrogenase (long-chain)	1.0	0.9	0.8	0.9	1.0	0.9	0.7	0.7	1.0	1.0	0.9	1.0	1.0	0.7	0.6	0.8
An17g01170	acyl-CoA dehydrogenase	1.0	1.1	1.1	1.1	1.0	1.2	0.9	1.1	1.0	0.9	0.9	1.0	1.0	1.2	1.0	1.2
An01g12960	*scdA* :: acyl-CoA dehydrogenase	1.0	1.1	1.3	1.7	1.0	1.0	1.7	**2.3**	1.0	1.3	1.4	1.6	1.0	**3.8**	**3.9**	**2.5**
An02g05840	enoyl CoA hydratase	1.0	1.0	1.1	1.1	1.0	1.2	1.2	1.0	1.0	1.2	1.1	1.1	1.0	1.7	1.8	1.2
An02g02820	*echA* :: enoyl-CoA hydratase	1.0	1.1	1.1	1.7	1.0	1.1	1.2	1.8	1.0	1.5	1.2	1.5	1.0	1.0	1.5	1.5
An14g00990	*foxA* :: trifunctional protein of β-oxidation	1.0	0.8	1.1	1.1	1.0	1.7	**3.5**	**4.6**	1.0	1.3	1.5	1.6	1.0	**8.7**	**8.6**	**4.1**
An14g00430	*hadA* :: 3-hydroxybutyryl-CoA dehydrogenase	1.0	1.2	1.4	1.8	1.0	1.1	1.8	**2.5**	1.0	1.2	1.2	1.6	1.0	**3.2**	**3.2**	**2.3**
An13g01920	*mthA* :: acetyl-CoA C-acetyltransferase	1.0	1.3	1.4	**2.0**	1.0	1.4	**2.2**	**3.1**	1.0	1.7	2.0	**2.2**	1.0	1.9	**3.5**	**2.4**
An04g05720	acetyl-CoA C-acyltransferase	1.0	1.2	1.3	1.6	1.0	1.4	**2.6**	**4.0**	1.0	1.1	1.4	1.5	1.0	**7.0**	**6.0**	**4.1**
An02g03320	acetyl-CoA C-acyltransferase (peroxisomal)	1.0	1.1	1.2	1.0	1.0	1.5	1.4	1.5	1.0	1.5	1.3	1.0	1.0	**3.2**	**2.8**	1.6
An08g05400	acetyl-CoA C-acyltransferase (peroxisomal)	1.0	1.2	1.3	1.4	1.0	1.4	2.0	**2.6**	1.0	1.1	1.3	1.5	1.0	**11.5**	**8.4**	**2.5**
An01g13490	acetyl-CoA C-acetyltransferase	1.0	1.1	1.2	1.2	1.0	1.7	1.7	2.0	1.0	1.2	1.3	1.1	1.0	**4.1**	**2.0**	1.6
	Auxillary enzymes																
An15g02860	2,4-dienoyl-CoA reductase (peroxisomal)	1.0	1.2	1.4	1.1	1.0	1.4	1.4	1.6	1.0	1.1	1.5	1.2	1.0	**3.7**	**3.7**	1.7
An09g02050	3-hydroxybutryl-CoA dehydratase	1.0	1.3	1.4	1.6	1.0	1.2	1.8	**2.4**	1.0	1.0	1.1	1.4	1.0	**3.1**	**4.0**	**2.9**
An15g01280	d3,d2-enoyl-CoA isomerase	1.0	0.9	1.1	1.0	1.0	1.0	1.4	1.2	1.0	1.2	1.0	1.0	1.0	**3.3**	**2.6**	1.0

Table 3 - β-Oxidation. This table presents the 19 genes that are predicted to be enzymes of the fatty acid degradation pathway. The bifunctional FoxA enzyme is active in two steps of this pathway. Transcript levels are relative to the signal of the non-induced sample taken from that fermentor (T=0). Bold font indicates over 2-fold expression, bold underlined font indicates over 3-fold expression.

are elevated already at 30 minutes after induction with the complex oil mixture. These four genes are all over 4-fold induced compared to non-induced levels, with An11g00400 induced almost 16-fold after 30 minutes.

The peroxisomal *foxA* ortholog, An14g00990, has elevated transcript levels up to 5-fold on olive oil and up to 9-fold up on wheat oil compared to non-induced levels (Table 3). The two genes putatively encoding enoyl-CoA hydratases, An02g02820 (the *echA* ortholog) and An02g05840, have expression levels that are below 2-fold increased both on olive oil and wheat oil.

The third step in β-oxidation in peroxisomes is also catalyzed by FoxA. In the mitochondrial pathway, An14g00430, the ortholog of *A. nidulans hadA*, follows the expression pattern of *foxA* in that it is up regulated on both olive oil and wheat oil. Expression is elevated to about 3-fold up on both oils.

The final step of β-oxidation is catalyzed by acyl-CoA C-acyltransferases, and all five predicted acyl-CoA C-acyltransferase encoding genes have elevated expression levels when induced with wheat oil. When induced with olive oil, only three acetyl-CoA C-acyltransferases are over 2-fold differentially expressed. The remaining two genes have increased transcript levels, but stay below 2-fold expression.

An13g01920 encodes the only acetyl-CoA C-acyltransferase that also has slightly elevated transcript levels after induction with DGDG or in the non-induced control, in an apparent time-dependent manner. This gene encodes the ortholog of the *A. nidulans mthA* gene.

Degradation of unsaturated fatty acids requires the recruitment of auxiliary enzymes. Odd-numbered double bonds are converted by Δ3,Δ2-enoyl-CoA isomerase and Δ3,5-Δ2,4-dienoyl-CoA isomerases, while even-numbered double bonds need the combined action of 2,4-dienoyl-CoA reductase preceding the isomerase [202, 244]. An15g02860 is predicted to encode the reductase, while the two isomerases are encoded by An15g01280 and An09g02050. These auxiliary enzymes have elevated transcript levels after induction with wheat oil. In the olive oil-induced fermentation, only isomerase An09g02050 has elevated transcript levels over 2-fold. This gene is also induced on pure DGDG and in the non-induced control fermentation, albeit mildly (1.6-fold).

When growing on acetate or other carbon sources that yield acetyl-CoA as sole carbon source, four-carbon molecules are required for energy production and biosynthesis. The glyoxylate shunt enables in two reaction steps the synthesis of the C4-compound succinate from two molecules of acetyl-CoA. Isocitrate lyase cleaves isocitrate to

succinate and glyoxylate. In the second reaction, glyoxylate and acetyl-CoA are condensed by malate synthase to malate. *AcuD* and *acuE*, the genes encoding isocitrate lyase and malate synthase respectively, have elevated expression levels in all four fermentations. Expression levels appear more elevated on wheat oil where *acuD* levels increase up to 8-fold at two hours after induction, while for the three remaining fermentations, expression levels are about 3-fold increased.

To summarize, genes encoding enzymes of all four steps of the β-oxidation pathway have elevated transcript levels when induced with olive oil and wheat oil. The spatial expression is different: on olive oil the transcript levels increase over time, while a maximum expression is reached after 30 minutes in the wheat oil induced fermentor. Both the peroxisomal and mitochondrial routes appear to be switched on, as both the peroxisome-localized FoxA and mitochondrial-localized EchA orthologs of *A. niger* are induced.

Transcriptional effects of the emulsifier Triton X-100

The emulsifier used in this study is octylphenol ethoxylate, or Triton X-100. This compound is able to permeabilize artificial membrane vesicles and is reported to affect membrane-associated enzyme activities [15]. Even though a minute amount of 0.002 % (v/v) Triton X-100 was added to the fermentations, a Triton-related transcriptional effect cannot be excluded. Fourteen genes are over 2-fold differentially expressed in the non-induced fermentation compared to the pre-induced time point. The expression pattern of these genes is summarized graphically in Figure 3. Remarkably, all 14 genes have an identical expression pattern for the olive oil-induced fermentor as well. The two fermentors induced with either DGDG or the complex wheat oil have under 2-fold elevated expression levels for these genes. The expression for these 14 genes is maximum at 30 minutes after induction, while decreasing to pre-induced levels after 2 hours.

Of the 14 genes, 4 genes encode a multidrug transporter, one gene encodes an acetoacetyl-CoA synthetase, and 6 genes encode proteins of unknown function. Two glutathione S-transferases and a glutathione peroxidase are encoded by the remainder of genes. These latter proteins relate to cellular protection, including oxidative stress and toxic foreign chemicals[102].

From these results, it is concluded that the fungus responds to the presence of Triton X-100 by co-expression of a relatively small number of dedicated genes only.

Uncharacterised transcriptional response

A diverse group of 129 genes has over 3-fold differentially expressed transcript levels compared to pre-induced levels in at least one fermentation sample. These genes and their expression levels are listed in Table 4. The expression profiles of these genes can be clustered in three groups.

A group of 51 genes has decreased transcript levels in the three oil-induced fermentations. Some but not all of these genes have also decreased expression levels on the non-induced fermentation.

A second group of 43 genes has increased transcript levels on both olive oil- and wheat oil- induced fermentations. These genes retain their transcript levels over the 2-hour period followed by microarrays. Expression of genes in this group is to a lesser extent also elevated on the DGDG-induced and Triton negative control fermentation samples. One-fifth of the genes in this group encode peptidases, and two peptide transporters are in this group.

Thirty-five genes form a third group of genes whose expression levels are elevated mainly at 30 minutes after induction with any of the used oil substrates. No elevated transcript levels are observed for the non-induced fermentation samples. A third of genes in this group encode proteins of unknown function. Three genes, encoding proteins similar to the fungus *Magnaporthe grisea* Pth11 membrane protein, are present in this list: An02g09500, An11g03730, and An14g05730. Pth11 family members are cell-surface G-protein-coupled receptors, and members of this protein family can bind exogenous as well as endogenous ligands, including lipids [147]. The Pth11 protein is required for pathogenicity of *M. grisea* [60]. Also, fatty acid β-oxidation is required for plant infection by this fungus [281]. It would be interesting to investigate the role of these Pth11-like proteins with regard to sensing and signalling of the presence of oil substrates in the *A. niger* environment.

Comparison with transcriptional response of *S. cerevisiae*

Peroxisome assembly and functioning has been studied by global transcriptional profiling in *S. cerevisiae* [237]. In that study, the yeast transcriptome was queried under two conditions that either induced or repressed peroxisome assembly and proliferation: by growth on oleate and glucose, respectively. A list of 224 candidate genes involved in peroxisome biogenesis or function was presented. For 111 of these listed genes an *A. niger* ortholog is present.

Transcript levels for this gene list were examined with the exception of An04g01535, for which no probe set is present on the DNA microarray. Thus, for 110 orthologs, transcript

levels were examined. Twenty-one genes are over 2-fold differentially expressed in at least one fermentation (Table 5). Apart from two hypothetical proteins for which no function is assigned all proteins are involved in core processes of fatty acid metabolism: peroxins, enzymes of the β-oxidation pathway, or fatty acid oxidation. The remaining 90 genes of this gene list have expression levels below 2-fold, and in most cases 1.4-fold or less. These genes are listed in Table 6.

Conclusion

The aim of this study was to determine whether our bioreactor-based experimental set up facilitates the study for global transcriptional responses towards lipids. Indeed, sorbitol-grown *A. niger* fermentor cultures respond to a pulse of 1 mM oil. This response is shown by elevated transcript levels of genes encoding enzymes of the fatty acid degradation pathway. Also, genes encoding peroxisome-proliferating proteins and transporters have elevated transcript levels after induction with oils.

This transcription response differs in magnitude between oils. While fermentor cultures that are induced with either olive oil or the complex wheat oil mixture show elevated expression of fatty acid related genes, almost no induction is detected after induction with digalactosyldiglycerides.

Next to the differing magnitude, the expression level profile for responding genes are different over time. For olive oil, transcript levels were elevated slightly after 30 minutes and continued to rise for 2 hours after induction. In contrast, in the wheat oil induced fermentation, expression levels were reached at 30 minutes followed by a decline of transcript levels. A major difference between the olive oil and wheat oil mixtures is the presence of free fatty acids in the latter, which add up to about half of this oil mixture. The readily available pool of fatty acids might be detected or metabolized earlier in time compared to olive oil, where free fatty acids first must be liberated through lipase activity.

One-hundred twenty-nine genes are over 3-fold differentially expressed in any of the four fermentations. Over a third of these genes (44 genes) encode proteins of unknown function, and also contains carbohydrate degradation and amino acid metabolism related proteins (Table 4). The majority of these genes is induced on the complex wheat oil mixture. Given its source as a crude extract from wheat gluten that likely contains many unidentified inducing compounds, this result is expected. However, this observation also suggests an overlap between metabolic and regulatory processes that under 'natural conditions' interact and might be interconnected.

Methods

Strain and spore preparation

A. niger strain 872.11 (Δ*argB*, *pyrA6*, *prtF28*, *goxC17*, *cspA1*) is derived from CBS 120.49. All media were based on minimal medium (MM) [204], contained 100 mM sorbitol as carbon source and were supplemented with uridine and arginine. To obtain spores, 20 spores per mm^2 were plated onto Complete Medium (CM) plates [204], incubated 5 days at 30 °C, and allowed to mature at 4 °C for 24 hours. The spore suspension was washed and stored at 4 °C until use.

Preparation of inducing emulsions

Pure wheat digalactosyldiglycerides (99%) were a kind gift of DSM Food Specialties (Delft, The Netherlands). Refined olive oil was purchased from Sigma. Wheat oil was extracted following [57] from wheat gluten obtained at the local mill. An 1:1.5 (w/v) ratio of gluten and ethanol was mechanically stirred for 1 hour, followed by filtration over a Büchner funnel with nylon filter (MW100 drd 15; Kabel Metaal). Another 1.5 volume of ethanol was added to the gluten fraction and processed alike. The ethanol was removed in a rotating evaporator under vacuum at < 40 °C. Twenty mg of this oil was dissolved in 1 mL $CDCl_3$/MeOD 1:1 (v/v), and ^{1}H-NMR spectra were recorded on a Bruker DRX-600 operating at a proton frequency of 600 MHz at a probe temperature of 300 K. ^{1}H-NRM analysis of the obtained oil shows that this oil fraction consists of 18% DGDG, 4.9% triglycerides, 47.7% free fatty acids and 4.0% mono-acyl DGDG. Both wheat MGDG and DGDG of wheat are assumed to be for over 80% in the 18:2 configuration [57].

To 20 mL of MM, 0.2% (v/v) Triton X-100 and 110 mM of each oil were added (which makes 1 mM final concentration after addition to the bioreactor). This solution was sonicated for 40 seconds (power '4', amplitude '20'). The negative control was sonicated as well.

Fermentation

Four glass 2.5-liter fermentors (Applikon) with 2.2 Liters of MM were kept at a constant temperature of 30 ± 0.5 °C while fermentor headplates were kept at 8 °C. 1.0×10^6 of spores per mL were added to the fermentor. During germination each fermentor was aerated through the headspace (50 L/h) and stirred at 300 rpm. When dissolved oxygen levels dropped below 60% for over 5 minutes the stirrer speed was set to 750 rpm and aeration was switched to the sparger inlet. This switching time point was defined as T=0 hours. Each fermentor was induced at T=14 hours. Twenty mL samples were removed from the fermentor and snap-frozen in liquid nitrogen directly after filtration.

RNA isolation and microarray processing

Frozen mycelium was ground using a dismembrator (Braun-Melsungen). A Trizol-chloroform extraction (Invitrogen) preceded total RNA extraction using RNeasy mini columns (Qiagen) according to the manufacturer's protocol for yeast. RNA integrity was assessed on an Experion system (Biorad). cDNA and cRNA synthesis and labeling, and array hybridization were performed following the Affymetrix users' manual [3] using the One-cycle Target Labeling and Control Reagents Kit to synthesize 15 µg of cRNA. Labeled cRNA was hybridized to Aspergillus niger GeneChip arrays at 45 ˚C for 16 hours. Washing and staining was done using the Hybridization, Wash and Stain Kit (Affymetrix). Arrays were scanned on an Agilent Technologies G2500A Gene Array Scanner (pixel value 3 µm, wave length 570 nm). The library and platform information for this proprietory dsmM_ANIGERa_coll511030F GeneChip (Affymetrix) is deposited in NCBI's GEO database (http://www.ncbi.nlm.nih.gov/geo/) under accession number GPL6758. The microarray data are submitted under accession number **GSE14285**.

Microarray data analysis

CEL files for individual array images were imported into GeneSpring 7.3 (Agilent technologies) using its RMA preprocessor [124]. Probe sets with values below 4 times the lowest probe signal on every array (50 AU) were removed, leaving 7,956 of 14,554 probe sets. Relative fold-change values were calculated per fermentation, by dividing each gene's raw signal value by that gene's value for the T=0 sample.

An in-house built database (Cluster of Orthologous Genes, or COG, database) [S. Basmagi and P. Schaap, unpublished data], based on publicly available yeast and fungal protein sequences, was used to transfer protein information between species. A protein clusters in a COG when corresponding ortholog sequences are detected in at least three organisms. The presence of a protein in such cluster is indicative of evolutionary conservation between species, and hints at protein functionality. Ortholog-base linking currently is the most powerful approach to transfer functional annotations and other related information between species [117]. Protein secretion signals were predicted by the SignalP program [187]. Two-thirds of the proteins in this database are highly conserved in the fungal kingdom, as orthologs are present in over 15 organisms. The COG database was used to transfer *S. cerevisiae* gene ontology information [109] to the orthologous *A. niger* genes. Thirty-six per cent (2,882 of 7,956) *A. niger* genes encodes a protein with *S. cerevisiae* ortholog in our database.

Acknowledgements

We would like to thank DSM Food Specialties for NMR-determination of the isolated wheat oil and for allowing the use of the *A. niger* DNA microarrays.

Gene name	Description	Group	DGDG				Olive oil				Non-induced				Wheat oil			
			0	0.5	1	2	0	0.5	1	2	0	0.5	1	2	0	0.5	1	2
An02g02930	str. sim. ribose-5-phosphate isomerase RpiB - *E. coli*	A	1.0	1.1	1.3	1.1	1.0	0.9	1.1	0.8	1.0	1.5	1.3	1.1	1.0	**0.3**	1.3	0.9
An02g08050	sim. phosphatidyl synthase - *S. pombe*	A	1.0	0.7	1.3	1.4	1.0	0.5	0.8	0.8	1.0	1.0	1.3	1.4	1.0	**0.3**	0.8	1.1
An08g11060	hypothetical protein	A	1.0	0.6	1.0	1.1	1.0	0.5	0.8	0.9	1.0	0.9	1.2	1.2	1.0	**0.3**	0.8	1.0
An07g04900	str. sim. mRNA sequence of cDNA clone 2589 - *A. niger*	A	1.0	0.8	1.0	1.2	1.0	0.6	0.8	0.9	1.0	1.1	1.1	1.0	1.0	**0.3**	0.8	1.0
An18g02680	str. sim. cytochrome P450 monooxygenase avnA - *A. parasiticus*	A	1.0	1.0	0.9	1.3	1.0	0.9	1.0	1.1	1.0	1.0	1.1	1.2	1.0	**0.3**	0.7	0.9
An08g10750	hypothetical protein	A	1.0	0.8	1.1	1.2	1.0	0.9	0.9	1.3	1.0	1.1	1.0	1.3	1.0	**0.2**	0.8	1.5
An09g00360	hypothetical protein	A	1.0	0.7	1.0	1.2	1.0	0.6	0.8	1.2	1.0	0.8	0.9	1.0	1.0	**0.3**	0.9	1.6
An01g08780	str. sim. high affinity hexose transporter Hxt1 - *S.cerevisiae*	A	1.0	0.9	1.0	1.1	1.0	0.6	1.1	1.1	1.0	0.9	1.1	1.2	1.0	**0.2**	1.5	1.2
An02g09310	str. sim. erythrocyte membrane protein stomatin - *H. sapiens*	A	1.0	0.9	1.2	1.8	1.0	0.7	1.1	1.5	1.0	0.8	1.2	1.4	1.0	**0.3**	0.9	1.4
An09g00270	alpha-galactosidase C aglC - *A. niger*	A	1.0	0.6	0.9	1.6	1.0	**0.4**	**0.4**	0.6	1.0	0.8	0.8	0.9	1.0	**0.3**	0.6	1.2
An08g03680	hypothetical protein	A	1.0	0.6	0.8	1.0	1.0	0.8	0.8	0.8	1.0	1.4	1.1	1.1	1.0	**0.3**	0.6	0.8
An05g01070	str. sim. 7-aminocholesterol resistance protein RTA1 - *S. cerevisiae*	A	1.0	0.7	0.8	0.9	1.0	0.9	0.7	0.6	1.0	1.7	1.1	1.1	1.0	**0.3**	0.6	0.6
An04g02970	sim. dimethylaniline monooxygenase FMO - *S. scrofa*	A	1.0	0.8	0.9	1.1	1.0	0.6	**0.5**	0.6	1.0	1.2	1.5	1.2	1.0	**0.3**	**0.5**	0.6
An15g04140	str. sim. polyketide synthase PKS1 - *C. heterostrophus*	A	1.0	0.7	0.8	0.9	1.0	0.6	**0.5**	0.7	1.0	1.1	1.5	1.3	1.0	**0.3**	**0.5**	0.7
An11g00040	hypothetical protein	A	1.0	0.7	0.7	0.9	1.0	0.5	**0.5**	0.5	1.0	1.4	1.2	1.1	1.0	**0.3**	**0.5**	0.7
An04g03840	sim. microtubule binding protein D-CLIP-190 - *D. melanogaster*	A	1.0	0.8	0.9	1.0	1.0	0.6	0.6	0.6	1.0	1.1	1.1	1.3	1.0	**0.3**	**0.5**	**0.4**
An01g08670	sim. nucleolin C23 - *C. griseus*	A	1.0	0.7	0.8	0.8	1.0	0.7	0.6	0.6	1.0	1.2	1.0	1.0	1.0	**0.3**	**0.5**	0.5
An13g03950	hypothetical protein	A	1.0	0.6	0.8	0.8	1.0	0.5	0.5	0.5	1.0	1.1	0.9	1.0	1.0	**0.3**	**0.4**	0.6
An15g02180	sim. salicylate hydroxylase nahW - *P. stutzeri*	A	1.0	0.9	0.8	1.0	1.0	0.7	**0.5**	0.9	1.0	0.7	0.9	1.2	1.0	**0.2**	**0.2**	**0.5**
An09g00220	str. sim. O-methyltransferase B (omtB) - *A. flavus*	A	1.0	0.7	0.9	1.2	1.0	0.7	0.6	0.8	1.0	1.0	1.0	1.3	1.0	**0.3**	**0.4**	0.8
An15g02190	str. sim. multidrug transporter bmr3 - *B. subtilis*	A	1.0	0.6	0.8	0.8	1.0	**0.4**	**0.5**	0.8	1.0	0.7	0.9	1.0	1.0	**0.3**	**0.3**	**0.5**
An13g01520	hypothetical protein	A	1.0	0.6	0.6	0.8	1.0	0.6	**0.3**	**0.4**	1.0	0.9	1.1	0.9	1.0	**0.4**	**0.3**	0.5
An02g10550	str. sim. endo-alpha-1 5-arabinanase precursor AbnA - *A. niger*	A	1.0	**0.3**	**0.4**	0.6	1.0	**0.3**	**0.3**	**0.4**	1.0	0.6	0.8	1.0	1.0	**0.3**	**0.3**	**0.4**
An02g07940	hypothetical protein	A	1.0	0.5	0.6	0.6	1.0	**0.5**	**0.4**	**0.4**	1.0	0.8	1.0	1.1	1.0	**0.4**	**0.3**	**0.4**
An07g04850	hypothetical protein	A	1.0	0.6	0.6	0.6	1.0	0.7	0.5	**0.4**	1.0	1.2	1.2	0.9	1.0	**0.5**	**0.3**	**0.3**
An07g02380	hypothetical protein	A	1.0	0.7	0.6	**0.5**	1.0	0.6	**0.4**	**0.5**	1.0	0.7	0.9	0.8	1.0	**0.4**	**0.3**	**0.4**
An02g07800	hypothetical protein	A	1.0	0.8	0.7	0.6	1.0	0.7	0.5	**0.4**	1.0	0.7	0.8	0.5	1.0	**0.3**	**0.3**	**0.4**
An02g03270	str. sim. methylenetetrahydrofolate reductase MET13 - *S. cerevisiae*	A	1.0	0.8	0.7	0.5	1.0	0.7	0.7	0.7	1.0	0.8	0.6	0.6	1.0	**0.3**	**0.3**	0.6
An15g02160	hypothetical protein	A	1.0	1.1	0.7	0.9	1.0	0.9	**0.4**	0.8	1.0	0.8	0.8	1.0	1.0	**0.5**	**0.3**	**0.5**
An01g09220	w. sim. tyrosinase melC2 - *S. lincolnensis*	A	1.0	0.9	0.9	0.9	1.0	0.7	0.6	0.7	1.0	0.8	1.0	1.0	1.0	0.6	**0.2**	0.6
An07g02360	sim. 6-hydroxy-D-nicotine oxidase 6-HDNO - *A. oxidans*	A	1.0	1.1	0.7	0.8	1.0	1.0	0.6	0.7	1.0	0.7	0.8	0.8	1.0	0.6	**0.3**	**0.5**
An12g04020	str. sim. acetyl-CoA carboxylase SPAC56E4.04c - *S. pombe*	A	1.0	0.9	0.8	0.9	1.0	1.0	0.7	0.7	1.0	0.8	0.9	0.9	1.0	0.6	**0.3**	**0.4**
An02g09530	sim. integral membrane protein Pth11 - *M. grisea*	A	1.0	1.3	1.1	1.0	1.0	0.7	0.6	0.5	1.0	0.6	0.9	0.8	1.0	**0.4**	**0.2**	**0.3**
An01g01120	hypothetical protein	A	1.0	0.9	0.9	1.0	1.0	0.6	**0.4**	0.5	1.0	0.7	0.9	0.8	1.0	**0.3**	**0.3**	**0.3**
An15g02200	str. sim. alcohol oxidase AOD1 - *C. boidinii*	A	1.0	1.0	0.8	0.9	1.0	0.7	0.5	0.6	1.0	0.9	0.9	1.0	1.0	0.6	**0.3**	**0.5**
An07g06820	str. sim. probable dehydrogenase - *Xanthobacter* sp.	A	1.0	0.8	0.8	0.9	1.0	0.7	**0.4**	**0.4**	1.0	0.8	1.0	1.1	1.0	**0.4**	**0.2**	**0.3**
An01g11320	sim. DOPA decarboxylase 3 - *A. oryzae*	A	1.0	0.7	0.6	0.6	1.0	0.7	**0.5**	**0.4**	1.0	0.7	0.7	0.6	1.0	0.6	**0.4**	**0.3**
An01g00060	str. sim. fatty acid synthase subunit alpha FAS2 - *S. pombe*	A	1.0	0.8	0.7	0.7	1.0	0.8	0.6	**0.5**	1.0	0.7	0.7	0.7	1.0	0.7	**0.3**	**0.3**
An08g05160	str. sim. oleate delta-12 desaturase odeA - *A. nidulans*	A	1.0	0.8	0.6	0.8	1.0	0.7	**0.5**	0.5	1.0	0.5	0.7	0.7	1.0	0.8	**0.3**	**0.3**
An11g00530	str. sim. ATP citrate lyase - *H. sapiens*	A	1.0	0.6	**0.4**	0.5	1.0	0.5	**0.5**	0.6	1.0	**0.5**	0.7	0.8	1.0	**0.3**	**0.1**	**0.2**

Gene name	Description	Group	DGDG				Olive oil				Non-induced				Wheat oil			
			0	0.5	1	2	0	0.5	1	2	0	0.5	1	2	0	0.5	1	2
An11g00510	str. sim. ATP citrate lyase ACL1 - *S. macrospora*	A	1.0	0.7	**0.4**	0.6	1.0	0.6	0.6	0.8	1.0	**0.5**	0.7	0.9	1.0	**0.3**	**0.2**	**0.2**
An15g07790	hypothetical protein	A	1.0	0.8	0.6	1.2	1.0	0.8	**0.4**	0.7	1.0	0.7	0.9	1.8	1.0	**0.4**	**0.2**	**0.4**
An08g11680	str. sim. 2,5-dichloro-2 (5-cyclohexadiene-1)4-diol dehydrogenase	A	1.0	0.9	0.7	1.3	1.0	0.6	**0.4**	0.6	1.0	1.1	1.4	1.4	1.0	**0.4**	**0.2**	**0.3**
An04g09990	str. sim. 2 5-dichloro-2 (5-cyclohexadiene-1) 4-diol dehydrogenase	A	1.0	0.6	0.7	0.7	1.0	0.8	**0.5**	0.7	1.0	1.0	1.3	1.5	1.0	0.5	**0.3**	**0.4**
An16g00220	hypothetical protein	A	1.0	0.7	0.9	0.7	1.0	1.0	**0.4**	**0.3**	1.0	**2.1**	0.8	**0.5**	1.0	**0.4**	0.6	0.7
An03g00690	hypothetical protein	A	1.0	1.0	1.0	0.5	1.0	1.0	**0.4**	**0.2**	1.0	1.2	1.0	**0.5**	1.0	0.7	**0.4**	**0.2**
An03g00680	str. sim. multidrug resistance protein FNX1 - *S. pombe*	A	1.0	0.8	0.9	**0.4**	1.0	0.7	**0.4**	**0.3**	1.0	0.7	0.7	0.5	1.0	0.5	**0.5**	**0.3**
An08g11120	hypothetical protein	A	1.0	0.8	0.6	0.6	1.0	0.7	**0.4**	**0.3**	1.0	1.0	0.9	0.6	1.0	0.6	**0.4**	**0.5**
An08g08840	str. sim. glutamate decarboxylase 1 GAD 1 - *A. thaliana*	A	1.0	0.7	0.6	**0.5**	1.0	0.8	**0.5**	**0.3**	1.0	1.2	1.0	0.5	1.0	1.1	0.9	0.6
An08g02300	w. sim. enniatin synthetase - *F. scirpi*	A	1.0	1.2	0.7	0.8	1.0	1.1	**0.3**	**0.3**	1.0	1.1	0.9	1.1	1.0	1.6	0.6	**0.5**
An14g01150	str. sim. glycine dehydrogenase - *P. sativum*	A	1.0	1.1	0.6	0.9	1.0	1.0	0.8	1.2	1.0	0.7	0.7	1.0	1.0	1.0	**0.3**	1.1
An11g03340	acid alpha-amylase - *A. niger*	B	1.0	1.2	1.4	**2.3**	1.0	1.0	1.6	**3.1**	1.0	0.9	0.7	1.0	1.0	1.0	**2.3**	**15.0**
An12g08280	exo-inulinase inu1 - *A. niger*	B	1.0	1.1	1.9	**8.5**	1.0	1.0	1.1	1.1	1.0	1.1	1.0	1.2	1.0	1.1	**4.7**	**9.9**
An14g04710	aspartic proteinase aspergillopepsin I pepA - *A. niger*	B	1.0	1.5	**2.7**	**3.0**	1.0	1.5	**4.9**	**9.2**	1.0	1.2	1.8	**2.3**	1.0	0.6	**5.7**	**5.2**
An06g02120	w. sim. Tyk2 non-receptor tyrosine - *M. musculus*	B	1.0	1.4	**2.9**	**3.5**	1.0	1.1	**3.2**	**5.9**	1.0	0.7	1.6	**2.3**	1.0	0.6	**5.9**	**5.2**
An08g04490	sim. putative serine peptidase - *O. sativa*	B	1.0	1.4	**3.2**	**2.4**	1.0	1.3	**4.4**	**4.8**	1.0	1.4	1.9	1.6	1.0	0.8	**7.1**	**5.0**
An15g07700	str. sim. aspergillopepsin II precursor - *A. niger*	B	1.0	1.3	1.7	1.6	1.0	1.1	1.9	**3.6**	1.0	0.9	1.1	1.1	1.0	1.1	**3.1**	2.0
An05g01860	hypothetical protein	B	1.0	1.3	1.4	1.8	1.0	0.9	1.7	**2.2**	1.0	1.2	1.4	1.5	1.0	0.8	**4.1**	**2.8**
An06g00190	str. sim. pepstatin insensitive protease CLN2 - *H. sapiens*	B	1.0	1.1	1.8	1.8	1.0	0.9	1.2	1.8	1.0	1.0	1.2	1.3	1.0	0.8	**3.1**	**2.4**
An04g03160	hypothetical protein	B	1.0	1.0	1.5	1.9	1.0	1.2	1.4	1.9	1.0	0.9	1.5	1.9	1.0	0.9	1.8	**3.2**
An08g09040	hypothetical protein	B	1.0	1.2	1.3	1.7	1.0	1.1	1.5	**2.1**	1.0	0.9	1.2	1.7	1.0	0.6	1.8	**3.1**
An01g05370	hypothetical protein	B	1.0	**3.4**	**2.6**	**3.7**	1.0	**3.4**	**4.7**	**7.0**	1.0	0.8	1.3	1.6	1.0	**2.5**	**4.9**	**5.9**
An01g07730	hypothetical protein	B	1.0	1.5	1.4	**2.5**	1.0	**2.8**	**2.2**	**5.0**	1.0	0.9	1.2	1.5	1.0	1.9	**2.6**	**4.9**
An09g02830	str. sim. acylaminoacyl-peptidase DPP V - *A. fumigatus*	B	1.0	1.7	1.6	1.7	1.0	1.7	**2.2**	**2.2**	1.0	1.1	1.3	1.3	1.0	1.2	**3.2**	**2.3**
An08g09420	str. sim. 4MeS - *M. anisopliae*	B	1.0	1.3	1.3	**2.9**	1.0	1.9	**2.6**	**10.9**	1.0	1.0	1.5	**2.7**	1.0	1.3	**3.2**	**9.4**
An03g06660	str. sim. peptide transporter ptr2 - *A. thaliana*	B	1.0	1.1	1.2	1.9	1.0	1.5	1.8	**3.2**	1.0	1.0	1.2	1.7	1.0	1.1	1.4	1.5
An12g03300	str. sim. aspartic protease pr1 - *P. rhodozyma*	B	1.0	1.4	1.9	1.0	1.0	1.4	**2.1**	1.6	1.0	1.2	1.1	0.8	1.0	1.2	**8.8**	**2.1**
An02g04690	str. sim. serine-type carboxypeptidase I cdpS - *A. saitoi*	B	1.0	1.8	**2.0**	1.2	1.0	**2.2**	**3.4**	**2.5**	1.0	**2.1**	1.6	1.1	1.0	1.2	**3.6**	1.8
An12g02660	hypothetical protein	B	1.0	1.6	1.4	1.4	1.0	1.6	1.7	1.7	1.0	1.1	0.8	1.0	1.0	**4.6**	**3.9**	**3.0**
An07g03290	sim. trans-2-enoyl-ACP reductase II fabK - *D. pneumoniae*	B	1.0	1.3	1.5	1.7	1.0	1.5	**2.5**	**3.3**	1.0	1.4	1.5	1.9	1.0	**5.4**	**6.7**	**4.7**
An16g05340	sim. trans-2-enoyl-ACP reductase II fabK - *D. pneumoniae*	B	1.0	1.4	1.5	1.5	1.0	1.5	**2.5**	**3.7**	1.0	1.3	1.4	1.5	1.0	**4.7**	**5.9**	**3.3**
An16g08040	str. sim. PMP24 protein - *M. musculus*	B	1.0	1.0	1.3	1.2	1.0	1.5	1.9	**2.2**	1.0	1.1	1.2	1.3	1.0	**4.4**	**3.7**	**2.2**
An16g06800	str. sim. endoglucanase eglB - *A. niger*	B	1.0	1.4	1.6	1.5	1.0	**2.0**	**2.0**	**2.2**	1.0	1.3	1.4	1.1	1.0	**4.2**	**3.4**	**2.2**
An13g01120	str. sim. leukotriene B4 12-hydroxydehydrogenase - *H. sapiens*	B	1.0	**2.5**	1.3	1.4	1.0	**5.1**	**5.8**	**11.5**	1.0	1.1	1.1	1.0	1.0	**43.0**	**15.0**	**4.8**
An08g04600	str. sim. peptide transporter Ptr2p - *S. cerevisiae*	B	1.0	1.7	1.4	1.2	1.0	**2.3**	**2.0**	**2.2**	1.0	1.0	1.1	1.1	1.0	**5.7**	**3.3**	1.7
An01g09310	sim. beta-ketoacyl reductase rhlG - *P. aeruginosa*	B	1.0	1.0	1.1	1.3	1.0	1.0	1.3	1.7	1.0	0.9	1.2	1.3	1.0	**5.0**	**4.2**	1.5
An08g08720	str. sim. cytochromeCc peroxidase precursor CCP1 - *S. cerevisiae*	B	1.0	1.0	1.0	1.0	1.0	1.3	1.5	**3.8**	1.0	1.2	1.1	1.1	1.0	**20.6**	**6.9**	**4.1**

Gene name	Description	Group	DGDG				Olive oil				Non-induced				Wheat oil			
			0	0.5	1	2	0	0.5	1	2	0	0.5	1	2	0	0.5	1	2
An08g05720	w. sim. soluble epoxide hydrolase (SEH) - *H. sapiens*	B	1.0	1.2	1.1	1.1	1.0	1.3	1.4	1.8	1.0	1.0	1.2	1.1	1.0	**4.7**	**3.4**	1.8
An11g05920	str. sim. prolidase - *A. esteraromaticum*	B	1.0	1.2	1.4	1.1	1.0	1.1	1.2	1.5	1.0	1.2	1.2	1.3	1.0	**3.3**	**2.2**	1.7
An02g03130	hypothetical protein	B	1.0	1.3	1.7	1.9	1.0	1.1	1.7	**2.9**	1.0	1.2	1.4	1.7	1.0	**2.4**	**2.7**	**3.1**
An15g01920	str. sim. methylcitrate synthase mcsA- *A. nidulans*	B	1.0	1.1	1.3	**2.3**	1.0	1.3	**2.2**	**3.9**	1.0	1.3	1.8	**2.4**	1.0	**2.6**	**3.4**	**2.9**
An01g08440	hypothetical protein	B	1.0	1.2	1.5	1.5	1.0	1.2	**2.2**	**3.2**	1.0	1.2	1.4	1.4	1.0	**2.0**	**4.4**	**3.9**
An16g07150	str. sim. cytoplasmic fumarate reductase FRDS1 - *S. cerevisiae*	B	1.0	1.0	1.3	1.3	1.0	1.1	**2.4**	**3.2**	1.0	1.0	1.3	1.3	1.0	**2.1**	**4.6**	**3.0**
An08g04640	str. sim. putative lysosomal protease CLN2 - *C. familiaris*	B	1.0	1.2	**2.7**	1.9	1.0	1.4	**3.0**	**4.0**	1.0	1.2	1.3	1.8	1.0	1.6	**7.1**	**5.0**
An12g01270	sim. hypothetical alanine-tRNA ligase alaS - *S. coelicolor*	B	1.0	1.2	1.4	1.3	1.0	1.0	1.4	1.4	1.0	1.3	1.4	1.6	1.0	**2.2**	**3.4**	1.9
An12g00170	str. sim. quinone reductase cryz - *C. porcellus*	B	1.0	1.0	1.4	1.2	1.0	1.0	1.8	**2.2**	1.0	1.0	1.0	1.5	1.0	**3.5**	**5.3**	**3.0**
An11g01400	hypothetical protein	B	1.0	1.1	1.5	1.6	1.0	1.1	1.7	**2.5**	1.0	1.3	1.4	1.5	1.0	**3.4**	**6.5**	**4.1**
An07g00130	hypothetical protein	B	1.0	1.0	1.4	**2.2**	1.0	0.9	1.2	1.7	1.0	1.2	1.3	1.9	1.0	**2.5**	**2.5**	**3.5**
An14g01110	str. sim. alkane-inducible cytochrome P450 gene - *Y. lipolytica*	B	1.0	0.9	0.9	1.1	1.0	1.0	1.2	1.7	1.0	1.4	1.3	1.4	1.0	**3.6**	**18.6**	**18.9**
An02g06440	str. sim. telomeric repeated gene RTM1 - *S. cerevisiae*	B	1.0	1.2	1.0	1.0	1.0	1.5	1.1	1.1	1.0	1.6	1.0	1.0	1.0	**6.6**	**7.6**	**8.4**
An11g03110	str. sim. methanol dehydrogenase Mdh - *B. methanolicus* C1	B	1.0	0.8	1.1	0.9	1.0	0.8	1.2	1.1	1.0	1.2	1.3	1.0	1.0	**3.0**	**3.3**	1.0
An06g02130	hypothetical protein	B	1.0	1.1	1.5	**3.0**	1.0	1.1	1.3	**3.0**	1.0	0.9	1.5	**2.1**	1.0	0.5	0.8	1.6
An18g02690	str. sim. dihydrogeodin oxidase DHGO - *A. terreus*	B	1.0	**2.3**	1.9	1.9	1.0	**2.5**	**2.5**	**4.2**	1.0	0.5	0.7	0.9	1.0	0.9	0.9	1.9
An07g00860	sim. 7-aminocholesterol resistance protein RTA1 - *S. cerevisiae*	B	1.0	1.7	**4.6**	**3.3**	1.0	1.0	1.1	1.1	1.0	1.0	1.0	1.1	1.0	0.9	1.0	1.0
An02g04900	endopolygalacturonases pgaB - *A. niger*	C	1.0	**2.4**	1.7	1.4	1.0	**4.2**	**2.4**	1.8	1.0	1.1	0.9	0.8	1.0	**4.7**	**3.7**	**2.3**
An11g03740	str. sim. isoamyl alcohol oxidase mreA - *A. oryzae*	C	1.0	**4.7**	1.3	1.1	1.0	**14.9**	**5.3**	**4.5**	1.0	1.2	1.0	1.0	1.0	**24.4**	**8.0**	**7.5**
An03g00640	sim. neutral amino acid permease mtr - *N. crassa*	C	1.0	**2.1**	1.5	1.2	1.0	1.9	1.8	1.8	1.0	1.0	0.9	0.8	1.0	**3.0**	**2.8**	1.7
An03g06270	str. sim. isoamyl alcohol oxidase mreA - *A. oryzae*	C	1.0	**2.5**	1.6	1.0	1.0	**4.3**	**2.8**	**2.8**	1.0	0.8	0.7	0.6	1.0	**3.4**	**2.2**	**2.3**
An12g04800	str. sim. adenosine deaminase AAH1 - *S. cerevisiae*	C	1.0	**2.5**	1.7	1.8	1.0	**2.9**	**2.1**	**2.8**	1.0	0.9	1.0	1.3	1.0	**3.8**	**2.5**	**2.9**
An14g05290	str. sim. protein isp4 - *S. pombe*	C	1.0	1.6	1.1	1.5	1.0	1.6	1.7	**2.2**	1.0	1.1	0.9	1.3	1.0	**3.2**	1.6	1.5
An01g07630	hypothetical protein	C	1.0	1.5	1.1	1.2	1.0	1.2	1.4	1.3	1.0	1.2	1.3	1.2	1.0	**4.7**	**2.1**	1.2
An13g02230	hypothetical protein	C	1.0	1.4	1.2	1.1	1.0	1.3	1.1	1.1	1.0	1.0	1.2	1.0	1.0	**3.8**	1.7	1.1
An15g00670	str. sim. alcohol acyl transferase EHT1 - *S. cerevisiae*	C	1.0	1.5	1.0	1.0	1.0	1.9	1.5	1.9	1.0	1.1	1.1	1.1	1.0	**6.2**	**2.3**	**2.1**
An14g03370	str. sim. allantoinase DAL1 - *S. cerevisiae*	C	1.0	1.2	1.0	1.1	1.0	1.6	1.2	1.3	1.0	1.3	1.1	1.1	1.0	**3.3**	1.5	1.4
An04g08990	hypothetical protein	C	1.0	1.2	1.0	1.0	1.0	1.7	1.3	1.3	1.0	1.0	0.9	0.9	1.0	**4.6**	1.6	1.2
An06g00280	hypothetical protein	C	1.0	1.4	0.9	1.0	1.0	1.8	1.1	1.3	1.0	1.0	1.0	1.0	1.0	**3.5**	1.4	1.5
An18g04430	sim. soluble epoxide hydrolase sEH clone EH3.1 - *S. tuberosum*	C	1.0	1.9	1.2	1.1	1.0	1.5	1.1	1.1	1.0	0.9	0.8	1.1	1.0	**4.4**	1.5	1.4
An08g06240	str. sim. uracil transport protein FUR4 - *S. pombe*	C	1.0	**2.2**	1.1	1.2	1.0	1.5	1.2	1.2	1.0	1.1	0.9	1.0	1.0	**4.0**	1.3	1.2
An02g09500	sim. integral membrane protein Pth11 - *M. grisea strain* G-11	C	1.0	1.6	1.1	0.8	1.0	**2.5**	1.4	1.2	1.0	1.1	1.0	0.9	1.0	**3.6**	1.2	1.1
An02g06090	w. sim. cellobiose dehydrogenase cdh - *T. versicolor*	C	1.0	1.6	1.1	0.8	1.0	**2.5**	1.7	1.6	1.0	0.9	1.0	0.8	1.0	**3.1**	1.5	1.3
An11g03730	w. sim. integral membrane protein PTH11 - *M. grisea* strain 4091-5-8	C	1.0	**2.0**	1.0	1.0	1.0	**5.3**	1.8	1.7	1.0	1.0	0.9	1.0	1.0	**7.1**	**3.0**	2.0
An09g05200	hypothetical protein	C	1.0	1.3	1.2	1.0	1.0	**2.2**	1.1	1.5	1.0	1.1	1.1	1.2	1.0	**3.0**	1.0	1.1
An16g02850	str. sim. cell wall glycosidase CRH1 - *S. cerevisiae*	C	1.0	1.3	1.4	1.2	1.0	1.6	1.5	1.6	1.0	1.1	1.3	1.0	1.0	**4.5**	1.1	1.5
An18g00680	hypothetical protein	C	1.0	1.2	1.1	0.5	1.0	1.7	1.0	0.8	1.0	1.0	1.0	0.8	1.0	**3.0**	1.3	1.0

Gene name	Description	Group	DGDG				Olive oil				Non-induced				Wheat oil			
			0	0.5	1	2	0	0.5	1	2	0	0.5	1	2	0	0.5	1	2
An12g03550	str. sim. allantoate permease DAL5 - *S. cerevisiae*	C	1.0	1.4	0.9	0.8	1.0	1.3	0.9	1.0	1.0	0.9	0.8	0.8	1.0	**3.6**	1.8	1.1
An02g13470	hypothetical protein	C	1.0	1.5	1.1	0.8	1.0	1.9	0.9	1.0	1.0	0.7	0.7	0.8	1.0	**3.8**	1.7	1.4
An14g05730	sim. integral membrane protein PTH11 - *M. grisea*	C	1.0	0.9	0.8	0.7	1.0	1.1	1.0	1.1	1.0	1.2	1.1	1.0	1.0	**3.2**	1.6	1.7
An16g02820	str. sim. fatty acid omega-hydroxylase CYP505 - *F. oxysporum*	C	1.0	0.9	1.1	1.0	1.0	1.6	1.2	1.5	1.0	**2.0**	1.8	**2.1**	1.0	**10.2**	1.8	1.4
An13g00140	str. sim. splicing factor YT521 - *R. norvegicus*	C	1.0	1.0	1.1	0.8	1.0	1.3	1.0	0.9	1.0	1.1	1.1	1.3	1.0	**3.7**	1.4	1.5
An14g02890	str. sim. 5-oxoprostaglandin 13-reductase LTB4/PGR - *S. scrofa*	C	1.0	1.1	1.1	1.0	1.0	1.2	1.0	1.0	1.0	0.9	1.0	1.0	1.0	**4.5**	1.7	1.0
An08g03040	hypothetical protein	C	1.0	0.9	0.9	0.9	1.0	1.3	1.1	1.2	1.0	0.9	1.0	0.9	1.0	**4.2**	1.5	1.2
An16g09050	str. sim. alcohol dehydrogenase alkJ - *P. oleovorans*	C	1.0	1.0	1.0	1.1	1.0	0.9	0.9	1.0	1.0	1.0	1.0	1.0	1.0	**9.6**	1.8	1.1
An12g09810	str. sim. thermostable alcohol dehydrogenase - *B. stearoth.*	C	1.0	1.0	1.0	1.0	1.0	1.0	1.1	1.1	1.0	1.1	1.0	1.0	1.0	**5.8**	1.2	1.1
An01g10840	w. sim. phosphoinositide 3-kinase adaptor BCAP - *M. musculus*	C	1.0	1.2	1.0	0.9	1.0	1.4	0.9	0.9	1.0	1.1	1.0	0.9	1.0	**3.1**	1.0	1.0
An09g03650	hypothetical protein	C	1.0	1.6	1.0	0.6	1.0	**2.9**	1.6	0.9	1.0	1.9	1.6	0.7	1.0	**3.5**	1.7	1.3
An02g11410	hypothetical protein	C	1.0	1.7	1.0	0.8	1.0	**2.5**	1.4	1.2	1.0	0.7	0.8	0.9	1.0	**3.3**	0.9	0.8
An02g11320	hypothetical protein	C	1.0	1.8	1.0	0.9	1.0	**2.0**	1.5	1.3	1.0	0.8	0.8	0.8	1.0	**3.5**	0.8	0.8
An18g01290	hypothetical protein	C	1.0	1.6	0.9	0.7	1.0	**3.5**	1.4	1.1	1.0	0.7	0.7	0.7	1.0	**2.3**	1.1	0.9
An17g02220	hypothetical protein	C	1.0	1.5	0.9	0.8	1.0	1.1	0.8	0.8	1.0	1.0	0.9	0.9	1.0	**4.1**	0.8	0.6

Table 4 (this page and previous pages) - A list of 129 genes that have relative expression levels over 3-fold compared to the non-induced fermentation sample (T=0) for at least one sampled condition. Bold font indicates over 2-fold expression, bold underlined font indicates over 3-fold expression.

Gene name	Yeast ID	Description	DGDG				Olive oil				Non-induced				Wheat oil			
			0	0.5	1	2	0	0.5	1	2	0	0.5	1	2	0	0.5	1	2
An16g04880	YKL197C	peroxin Pex1	1.0	1.2	1.1	1.1	1.0	1.2	1.3	1.3	1.0	1.0	1.0	1.0	1.0	**2.2**	1.5	1.1
An16g02010	YNL329C	peroxin Pex6	1.0	1.1	1.2	1.0	1.0	1.1	1.3	1.3	1.0	1.3	1.2	1.1	1.0	**3.4**	**2.1**	1.1
An16g07610	YLR191W	peroxin Pex13	1.0	1.0	1.1	1.1	1.0	1.3	1.4	1.6	1.0	0.9	1.0	1.0	1.0	**2.0**	1.9	1.3
An01g03680	YPL147W	peroxisomal ABC transporter (Pxa1)	1.0	0.9	1.2	1.1	1.0	1.1	1.9	**2.3**	1.0	1.1	1.4	1.5	1.0	1.5	**2.9**	**2.1**
An09g06740	YER015W	long chain fatty-acid-CoA ligase (Faa2)	1.0	**2.1**	1.3	1.3	1.0	**3.2**	2.0	**2.2**	1.0	1.2	1.3	1.0	1.0	**8.6**	**2.6**	1.4
An03g03360	YOR100C	carnitine acetyltransferase (AcuH, *A. nidulans*)	1.0	1.4	1.5	1.4	1.0	1.3	2.0	**2.3**	1.0	1.3	1.4	1.3	1.0	**5.5**	**4.4**	2.0
An18g01590	YML042W	carnitine acetyltransferase	1.0	1.4	1.5	1.5	1.0	1.8	**3.0**	**3.4**	1.0	1.2	1.3	1.2	1.0	**8.4**	**5.1**	**2.2**
An14g00990	YKR009C	FoxA, trifunctional protein ofthe beta-oxidation	1.0	0.8	1.1	1.1	1.0	1.7	**3.5**	**4.6**	1.0	1.3	1.5	1.6	1.0	**8.7**	**8.6**	**4.1**
An04g05720	YIL160C	acetyl-CoA acyltransferase	1.0	1.2	1.3	1.6	1.0	1.8	**3.4**	**5.1**	1.0	1.1	1.4	1.5	1.0	**7.0**	**6.0**	**4.1**
An01g13490	YJR019C	acetyl-CoA acyltransferase	1.0	1.1	1.2	1.2	1.0	1.4	1.5	1.7	1.0	1.2	1.3	1.1	1.0	**4.1**	**2.0**	1.6
An15g02860	YNL202W	peroxisomal 2,4-dienoyl-CoA reductase	1.0	1.2	1.4	1.1	1.0	1.3	1.3	1.5	1.0	1.1	1.5	1.2	1.0	**3.7**	**3.7**	1.7
An15g01280	YLR284C	d3,d2-enoyl-CoA isomerase	1.0	0.9	1.1	1.0	1.0	0.9	1.2	1.1	1.0	1.2	1.0	1.0	1.0	**3.3**	**2.6**	1.0
An01g09270	YER065C	isocitrate lyase	1.0	1.0	1.7	**3.2**	1.0	1.1	**2.4**	**3.9**	1.0	1.1	1.7	**2.8**	1.0	**2.4**	**5.5**	**7.9**
An15g01860	YNL117W	malate synthase	1.0	1.1	0.8	1.7	1.0	1.0	1.7	**4.4**	1.0	1.1	1.1	**2.1**	1.0	**3.4**	**2.4**	**3.8**
An12g07630	YPR006C	2-methylisocitrate lyase (Icl2)	1.0	1.3	1.4	1.7	1.0	1.2	1.5	**2.2**	1.0	1.5	1.6	1.9	1.0	1.9	**2.5**	**2.1**
An04g09030	YJR095W	mitochondrial succinate-fumarate transporter (Sfc1)	1.0	1.1	1.1	1.5	1.0	1.1	1.6	**2.5**	1.0	0.9	1.0	1.1	1.0	2.0	**2.2**	2.0
An07g08280	YML075C	3-hydroxy-3-methylglutaryl-CoA reductase (Hmg1)	1.0	1.2	0.9	0.8	1.0	1.3	1.0	1.0	1.0	0.8	0.9	0.9	1.0	**2.1**	1.1	1.0
An08g05780	YKL188C	peroxisomal ABC transporter (Pxa2)	1.0	1.2	1.1	1.2	1.0	1.1	1.9	1.9	1.0	1.0	1.0	1.2	1.0	1.5	**2.3**	**2.1**
An15g00670	YMR210W	putative acyltransferase	1.0	1.5	1.0	1.0	1.0	1.9	1.5	1.9	1.0	1.1	1.1	1.1	1.0	**6.2**	**2.3**	**2.1**
An08g03040	YBL095W	putative protein of unknown function	1.0	0.9	0.9	0.9	1.0	1.3	1.1	1.2	1.0	0.9	1.0	0.9	1.0	**4.2**	1.5	1.2
An01g07630	YOR084W	peroxisomal matrix localized lipase (Lpx1)	1.0	1.5	1.1	1.2	1.0	1.2	1.4	1.3	1.0	1.2	1.3	1.2	1.0	**4.7**	**2.1**	1.2

Table 5 - 21 of 111 orthologs of *S. cerevisiae* oleate-induced genes [237] that have relative expression levels over 2-fold for at least one microarray. Bold font indicates over 2-fold expression, bold underlined font indicates over 3-fold expression.

Gene name	Yeast ID	Gene name	Yeast ID	Gene name	Yeast ID
An01g00110	YHR202W	An06g01080	YDR205W	An13g00660	YMR041C
An01g04450	YOR360C	An07g05930	YDL219W	An13g01090	YJR098C
An01g04630	YBR039W	An07g07000	YMR089C	An14g01500	YGL196W
An01g05330	YBR170C	An07g07100	YIL064W	An14g01590	YML008C
An01g06180	YOR065W	An07g08890	YGR086C	An14g01700	YML123C
An01g06380	YHR140W	An08g00210	YIL155C	An14g02230	YMR216C
An01g09050	YAL053W	An08g00450	YHR132W-A	An14g06000	YJL112W
An01g10170	YLL019C	An08g01000	YFL030W	An14g06290	YDR036C
An01g11720	YPL207W	An08g01600	YBL008W	An14g06530	YJL039C
An01g12590	YKL137W	An08g03340	YOR043W	An15g00070	YDL078C
An01g14790	YJL218W	An08g04300	YJL084C	An15g00170	YMR139W
An02g02750	YDR256C	An08g04990	YAR035W	An15g00410	YPL276W
An02g07500	YIR034C	An08g06370	YBL033C	An15g01710	YKL016C
An02g07530	YKL054C	An08g06580	YMR280C	An15g02340	YOL058W
An02g09350	YPR003C	An08g06850	YDL020C	An15g02490	YIL094C
An02g10780	YDL167C	An08g07470	YLR093C	An15g04490	YMR109W
An02g11720	YGL156W	An09g04520	YLR285W	An15g04520	YMR001C
An02g12170	YOR188W	An10g00530	YJL153C	An15g06030	YKR076W
An03g06880	YBL017C	An11g00470	YOR095C	An15g06820	YPR117W
An03g06950	YLL001W	An11g02590	YOL147C	An16g02150	YNL212W
An04g01200	YDR529C	An11g09400	YDR377W	An16g04020	YGL153W
An04g01535	YKL087C	An11g09870	YLR246W	An16g04500	YDR247W
An04g04130	YLR438W	An11g11240	YGL143C	An16g07290	YPL078C
An04g05260	YNR050C	An12g03940	YMR095C	An16g07480	YDR244W
An04g05270	YDR214W	An12g06300	YNL134C	An16g07490	YEL020C
An04g05620	YLR153C	An12g07510	YGR243W	An18g02830	YOL087C
An04g05630	YGL092W	An12g08040	YIR015W	An18g03170	YPL110C
An04g06040	YPR023C	An12g08690	YOR370C	An18g04220	YBL030C
An04g08760	YHR170W	An12g08730	YMR043W	An18g04540	YBR046C
An05g00110	YNR020C	An12g08800	YLR430W	An18g04590	YDL135C

Table 6 - Orthologs of *S. cerevisiae* oleate-induced genes. A list of 90 of 111 genes for which their *S. cerevisiae* ortholog is differentially expressed under peroxisome proliferating conditions [237], but that have expression levels below 2-fold on all conditions in this study. Yeast ID, *S. cerevisiae* locus tag.

Chapter 5

Gene co-expression networks in *Aspergillus niger*

Douwe van der Veen[‡], Machtelt Braaksma[‡], Robert A. van den Berg[‡],

Mariët J. van der Werf, Peter J. Punt,
John van der Oost, and Leo H. de Graaff

[‡] These authors contributed equally to this work

Abstract

The fungus *Aspergillus niger* has been studied in considerable detail with respect to various industrial applications. Still, the mechanisms that control the adaptation of its physiology are poorly understood. In this study, clustering of co-expressed genes has been performed on the basis of DNA microarray data sets of two experiments in which *A. niger* was exposed to either a relatively mild or a relatively severe physiological stress.

Three gene co-expression networks were constructed for both individual data sets as well as for the combined data. These networks are composed of modules that are enriched for genes that encode proteins involved in similar biological processes, including ribosome biogenesis, fatty acid metabolism, and amino acid metabolism. Conserved sequences are detected in the upstream region of genes that cluster in some modules, such as the motif for the amino acid metabolism-related transcription factor CpcA and the motif for the fatty acid metabolism-related transcription factors, FarA and FarB.

Comparative analysis revealed that some of the identified modules are present in all three networks. However, for each network, different co-expressed gene pairs connect the modules. This observation would be in line with gene expression being co-ordinated at different hierarchical levels.

Introduction

Soil is the natural habitat of *Aspergillus niger*, where this fungus feeds on decomposing plant material. It survives in this harsh environment by its ability to secrete a range of macromolecule-degrading enzymes that release nutrients and by its ability to inhibit the growth of competitors by acidification of its surroundings. In industry, these capacities are well exploited: *A. niger* is an efficient producer of both homologous and heterologous proteins [196, 207] and is the major production organism for citric acid world-wide [167].

A. niger has the potential to adjust to a wide range of cultivation conditions; however, the underlying regulatory mechanisms are poorly understood. The availability of annotated genome sequences of *A. niger* [128, 196] as well as technologies built upon these genome sequences such as DNA microarrays, now allow for in-depth studies to increase our understanding of these mechanisms.

With DNA microarray technology, snapshots of global transcription levels can be taken from varying experimental conditions. The thus obtained gene transcription data can be considered as a reflection of the physiological state of *A. niger* and can be interpreted with the aim to understand its behavior. *A. niger* transcriptome studies have been used to characterize polysaccharide-degrading enzyme systems [11, 130, 173, 264, 294, 295], to study the *A. niger* response towards reductive stress [91] or cell wall damage [176], and to describe spatial colony development [153].

In the above-mentioned transcriptome studies, emphasis is placed on the comparison between two or more experimental conditions. These studies focus on the specific differences in gene transcription levels between experimental conditions. It is often assumed that the genes identified in such studies are involved in the direct response to the experimentally applied stimuli. However, this is likely an oversimplification of the processes leading to the observed differential expression, because modulated gene expression may frequently be triggered indirectly. An analysis of microarray data of 276 *Saccharomyces cerevisiae* deletion strains showed that, for over 95% of these deletion strains, at least one additional gene has a highly significant altered expression profile [116]. Featherstone and Brodie deduced from these microarray data (i) that specific sets of genes with similar expression profile are present, (ii) that such interactions are widespread at the level of gene expression, and (iii) that these interactions can be described within a gene co-expression network [77]. Such networks can provide a large-scale, global view of the physiological state of an organism at the genetic level.

A comparison of gene co-expression networks constructed from microarray data of six evolutionary distinct organisms including yeast, *Escherichia coli*, and *Arabidopsis thaliana*,

indicated that these networks share common topological properties [25]. For example, most genes of such a network have an expression profile that correlates with that of few other genes, while few genes have an expression profile that correlates with that of many genes. These networks are generally robust: given that most genes partner with few other genes based on their expression profiles, removal of a gene from the network will in most times not affect the overall topology of the network [7, 17]. For example, a disruption in one of the *A. niger* extracellular xylanase-encoding genes does not likely impair this fungus' ability to degrade xylan, whereas a mutation in the transcription factor of the xylanolytic system greatly influences its xylan-degrading ability. In addition to the robustness of such networks, the observed gene co-expression networks are characterized by groups of highly interconnected groups of genes, called clusters or modules. Such modules can be, but do not have to be, working together to achieve a distinct biological function [71].

However, the aforementioned study by Bergmann and co-workers also indicated that the relations between the functional modules vary significantly between the six organisms [25]. For example, the average gene expression profiles for genes in the 'secreted protein' and 'proteasome' modules correlate positively in yeast and *A. thaliana*, negatively in *Drosophila melanogaster*, and do not appear to correlate significantly in *Homo sapiens*. These different relations between modules within these organisms might be explained by the presence of organism-specific regulatory circuits. Already within fungi, the low evolutionary conservation of *cis*-regulatory elements suggests a rather high level of regulatory diversification. Gasch and co-workers investigated evolutionary conservation and divergence for 42 *S. cerevisiae* regulons in 14 ascomycete fungi [84]. For example, the 5'-TGACTM *cis*-regulatory element is present in the upstream region of genes encoding amino acid biosynthesis enzymes of 12 out of 14 fungi. However, since only 3 of 42 regulatory modules were found to be conserved between *S. cerevisiae* and *A. nidulans*, motif conservation appears to be the exception rather than the rule [84].

The above-mentioned studies point out that information of gene co-expression networks of other organisms including *S. cerevisiae* cannot be directly translated to the global patterns underlying gene transcription in *A. niger*. In this study, we aimed to obtain an initial view on the generic organization of gene expression profiles of *A. niger*. This study is restricted to global biological differences because rather different experimental conditions were used.

For our analysis, we examined gene transcription profiles derived from two microarray data sets. The first set of microarrays is obtained from identical bioreactor-cultivated fungal cell cultures that have been pulsed with low amounts of various inducers. These inducing compounds include the sugar D-xylose as well as three lipids. We reasoned that

these pulses of low amounts of inducer will provoke only a minor disturbance (mild stress) of the gene co-expression network and that transcription profiles will only change due to the inducing effect of these compounds.

The second set of microarray data is derived from bioreactor-cultivated cells that were grown on different media for which factors like carbon source and pH were varied. For this data set, more drastic changes (severe stress) in transcription profiles are expected, as the fungus will need to accommodate to these different conditions.

Description of global patterns of gene transcription is still a largely unexplored area of investigation. This study is an initial attempt to investigate such global patterns in *A. niger*, and aims to contribute to a better understanding of its physiology.

Results

A subset of genes is selected for analysis

The *A. niger* strain CBS 513.88 genome sequence contains over 14 thousand predicted protein-encoding genes. Due to computational constraints, only a subset of all predicted genes could be used for this analysis. In this study, we chose to examine only those protein-encoding genes for which an ortholog is present in 15 or more fungal species (see Materials and Methods). Manual annotation of the genome notwithstanding, for a large proportion of proteins their physiological function is unknown, while many of the proteins whose biological process or molecular function are 'known' are poorly understood [115]. We considered that a protein's evolutionary conservation suggests a functional role, even in case no clear biological function has been assigned. A drawback of this selection criterion is the exclusion of most genes encoding biopolymer-degrading hydrolases and other extracellular enzymes, as these enzymes are generally less conserved. Because of the excellent annotation information that is available for *S. cerevisiae* proteins, all *A. niger* genes that are orthologs for proteins encoded by this model organism were also added.

Not all of the *A. niger* genes are expressed within the two sets of conditions. Under mildly perturbed conditions, 6,599 genes (45%) are evaluated as present on microarrays of this data set. On strongly perturbed conditions microarrays, 9,470 genes (65%) are evaluated as present. Our working gene list comprised 2,773 conserved protein-encoding genes that are present on the majority of microarrays obtained under both mildly and strongly perturbed conditions. These genes make up for 42% and 29% of the expressed genes within the mildly and strongly perturbed conditions data set, respectively.

The transcript level data for these 2,773 genes were used to construct three gene co-expression networks: one network for each of the two sets of experimental conditions, and one network constructed from transcript level data of the combined data sets.

Gene co-expression networks are constructed

The gene co-expression correlation coefficient ρ was calculated for all pair-wise combinations of the 2,773 genes for each of the three data sets. The distribution of ρ within each data set is given in Figure 1 A. The histogram for the mildly perturbed data set indicates that many gene pairs correlate relatively poorly (Figure 1 A, left). Under strongly perturbed experimental conditions, genes tend to be more strongly correlated or anti-correlated as can be seen by the broader base of the histogram (Figure 1 A, middle).

The calculated ρ values were used to construct three gene co-expression networks. In such network, a connecting line is drawn between each pair of genes for which their

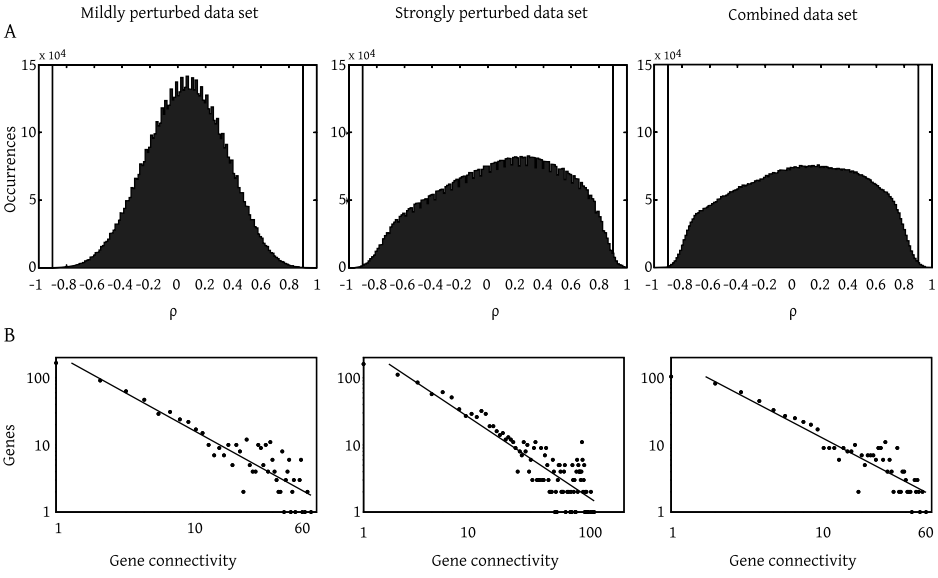


Figure 1 - (A) Histogram of the distribution of ρ as calculated for all possible gene pair combinations within the subset of 2,773 genes. The distribution of ρ is obtained by dividing the total range of ρ in equally spaced bins, followed by counting how often a value for ρ for a gene pair is found per bin. (B) For each gene, the number of gene pairs it partners with (connectivity; horizontal axis) is plotted against the number of genes with identical number of gene pairs. The fitted line is for a power-law distribution, $P(k) \sim k^{-\gamma}$. For all networks, γ is around 1.2.

expression profiles correlate stronger than a given threshold. For this study, the threshold for strongly co-expressed expression profiles was set at an absolute ρ of 0.90 or higher. At lower values of ρ, the network of the strongly perturbed conditions data becomes too cluttered by connected gene pairs, which renders effective comparison of the three networks difficult. We observe for all three networks that there are many genes with few connections to other genes (*i.e.* these genes have a low connectivity). In contrast, the expression profile for few genes correlates with that of many other genes (Figure 1 B).

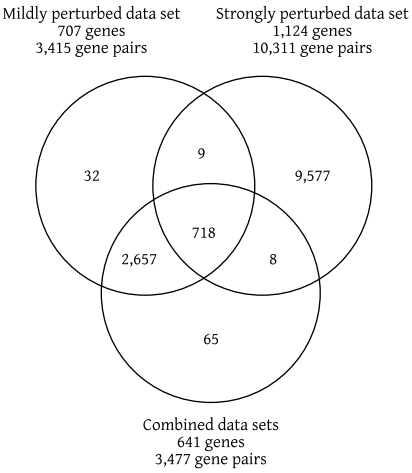


Figure 2 - Venn diagram indicating the number of correlating gene pairs between the three data sets. All gene pairs have a ρ equal to or higher than 0.90.

We examined the extent of overlap between the most strongly co-expressed gene pairs within the three constructed networks (Figure 2). The Venn-diagram indicates that the cross-section of all three networks harbors 718 pairs of strongly co-expressed genes. These 718 gene pairs are formed between 194 genes (Table 1). In addition to these 194 genes, a further 167 genes are present in all three data sets. For these 167 genes, their expression profile correlates strongly with that of at least one other gene for all three networks; however, they are partner in different gene pairs within each network (*e.g.* when imagining a network X-Y-Z and a second network X-Z-Y, all three genes (X, Y, and Z) are in both networks, but only one gene pair (Y-Z) is present in both networks).

The combined 194 and 167 genes take part in a disproportionate large number of strongly co-expressed gene pairs relative to the whole network. For example, for the mildly perturbed conditions network, these genes correspond to half of the total number of co-expressed genes (361 of 707 genes) but are connected to 80% of the gene pairs in this network. This observation indicates that these 361 genes have a higher

Table 1 (next two pages) – Genes within each of the 9 individual modules observed within the network that was constructed from the 718 gene pairs present in all three data sets at an absolute ρ value greater than or equal to 0.90 are listed. The module labels A through I correspond to the coloring scheme used in Figures 3 and 4. Abbreviation (s.)sim. is for "(strong) similarity to". # Due to space constraints, protein-encoding genes with highly similar annotations were grouped.

Gene name	Description
	module A
An02g07210	aspartic protease pepE
An05g00080	sim. membrane protein CBP3 - *S. cerevisiae*
An12g07790	s.sim. transcription factor BTF3 - *S. pombe*
8 genes #	s.sim. cytoplasmic acidic ribosomal protein
33 genes #	cytoplasmic ribosomal protein of the large subunit
23 genes #	cytoplasmic ribosomal protein of the small subunit
An13g00080	s.sim. conserved hypothetical protein B2F7.100 - *N. crassa*
An18g02890	s.sim. mitochondrial ribosomal protein L11 - *S. cerevisiae*
An18g05630	sim. mitochondrial ribosomal protein S2 MRP4 - *S. cerevisiae*
An08g04070	s.sim. mitochondrial receptor complex chain MOM22 - *N. crassa*
An02g04260	s.sim. small hypothetical zinc-finger protein TIM13 - *S. pombe*
An04g05950	s.sim. mitochondrial ribosomal protein L24 MRPL24 - *S. cerevisiae*
An15g01700	s.sim. nascent polypeptide-associated complex alpha chain - *M. musculus*
	module B
An14g01560	s.sim. zuotin ZUO1 - *S. cerevisiae*
7 genes #	s.sim.translation initiation factor
An04g00680	sim. nucleolar protein gar2 - *S. pombe*
An07g08170	s.sim. IMP dehydrogenase IMH3 - *C. albicans*
An01g08850	s.sim. Gbeta like protein CpcB - *A. nidulans*
An01g12230	s.sim. nucleolar protein NOP1 - *S. cerevisiae*
An12g00470	sim. hexamer-binding protein HEXBP – *L. major*
An01g09100	s.sim. actin-related protein ArpA - *A. oryzae*
An11g05510	s.sim. peptidylprolyl isomerase FPR3 - *S. cerevisiae*
An15g00190	s.sim. mitochondrial import receptor MOM38 - *N. crassa*
An08g04470	s.sim. mitochondrial elongation factor Tu - *A. thaliana*
An07g06520	s.sim. heat shock protein 70 H.log pdr13 - *S. cerevisiae*
An16g09260	s.sim. dnaK-type molecular chaperone SSB2 - *S. cerevisiae*
An17g02170	s.sim. centromere/microtubule-binding protein CBF5 - *S. cerevisiae*
An02g12750	s.sim. cytoskeleton specific chaperonin subunit CCT4 - *S. cerevisiae*
An08g07050	s.sim. 189 kD subunit of DNA-directed RNA polymerase I RPA190 - *S. pombe*
An04g02000	s.sim. hypothetical microtubule-interacting protein YTM1 - *S. cerevisiae*
An04g02550	s.sim. mitochondrial translation elongation factor EF-G MEF1 - *S. cerevisiae*
	module C
An02g01830	cytochrome c cyc - *A. niger*
An14g04170	cytochrome c oxidase subunit V cox5 - *A. niger*
An08g01370	s.sim. Oxodicarboxylate carrier ODC2 - *S. cerevisiae*
An11g05700	sim. probable membrane protein YGR235c - *S. cerevisiae*
An02g09930	s.sim. subunit VI of cytochrome c oxidase COX6 - *S. cerevisiae*
An07g07390	s.sim. subunit IV of cytochrome c oxidase COX4 - *S. cerevisiae*
An04g01200	s.sim. 14 kD subunit of ubiquinol--cytochrome c reductase Qcr7 - *S. cerevisiae*
	module D
An16g03040	s.sim. AMMECR1 - *H. sapiens*
An02g14910	sim. human peptidase HPEP-14 from patent WO200042201-A2
An11g11320	s.sim. protein involved in autophagy Aut2p - *S. cerevisiae*
An02g04850	s.sim. the hypothetical protein encoded by An07g08350 - *A. niger*
	module I
An11g11300	histone H2A httA - *A. niger*
An11g11310	s.sim. histone H2B - *A. nidulans*
An08g06940	s.sim. histone H4.1 - *A. nidulans*

Gene name	Description
	module E
An08g05300	s.sim. heat shock protein hsp70 pss1 - *S. pombe*
An11g09790	s.sim. sulfate adenylyltransferase sC - *A. nidulans*
An16g05090	s.sim. endonuclease SceI 75 kDa subunit ENS1 - *S. cerevisiae*
An12g04940	s.sim. mitochondrial heat shock protein HSP60 - *S. cerevisiae*
An04g07330	s.sim. the mitochondrial ribosomal protein MRPL12 - *H. sapiens*
An07g09530	s.sim.subunit pyruvate dehydrogenase complex PDA1 - *S. cerevisiae*
An11g11280	s.sim. the dihydrolipoamide succinyltransferase KGD2 - *A. fumigatus*
An01g00100	s.sim. pyruvate dehydrogenase beta chain precursor PDB1 - *S. cerevisiae*
An08g05580	s.sim. isocitrate dehydrogenase (NAD+) chain IDH2 precursor - *S. cerevisiae*
	module F
An15g01920	s.sim. methylcitrate synthase mcsA- *A. nidulans*
An12g07630	s.sim. 2-methylisocitrate lyase ICL2 - *S. cerevisiae*
An15g01280	s.sim. d3 d2-Enoyl-CoA Isomerase Eci1 - *S. cerevisiae*
An01g09830	s.sim. glutathione S-transferase GTT1 - *S. cerevisiae*
An04g05720	s.sim. acetyl-CoA C-acyltransferase POT1 – *Y. lipolytica*
An16g05340	sim. trans-2-enoyl-ACP reductase II fabK – *S. pneumoniae*
An13g01920	s.sim. acetyl-CoA C-acetyltransferase precursor – *R. norvegicus*
An16g07150	s.sim. soluble cytoplasmic fumarate reductase FRDS1 - *S. cerevisiae*
An14g00990	s.sim. trifunctional protein of the beta-oxidation fox-2 - *N. crassa*
	module G
An04g02010	sim. ribosomal protein S13 rpsM – *B. subtilis*
An18g04490	sim. the 30s ribosomal protein s19 RpsS – *E. coli*
An12g03890	s.sim. mitochondrial ribosomal protein L14 - *S. cerevisiae*
An12g00110	s.sim. F1-F0 complex assembling factor atp12 - *S. cerevisiae*
An07g06570	s.sim. mitochondrial ribosomal protein S14 mrp2 - *S. cerevisiae*
An02g03110	sim. mitochondrial ribosomal protein of the large subunit YmL4 - *S. cerevisiae*
	module H
An02g14380	hexokinase hxk - *A. niger*
An11g10150	s.sim. SAICAR synthase ADE1 - *S. cerevisiae*
An17g02280	s.sim. aspartate kinase HOM3 - *S. cerevisiae*
An04g01340	s.sim. asparagine synthase Asn1 - *S. cerevisiae*
An16g02520	s.sim. threonine synthase (THR4) - *S. cerevisiae*
An01g13920	s.sim. adenylosuccinate synthase ade2 - *S. pombe*
An04g02580	s.sim. threonine deaminase ILV – *A. adeninivorans*
An01g06560	s.sim. argininosuccinate lyase ASAL - *C. albicans*
An01g08490	s.sim. histidine--tRNA ligase HTS1 - *S. cerevisiae*
An04g05260	s.sim. the saccharopine reductase LYS3 – *M. grisea*
An08g02990	s.sim. adenylylsulfate kinase MET14 - *S. cerevisiae*
An02g07500	s.sim. saccharopine dehydrogenase LYS1 - *C. albicans*
An13g01080	s.sim. ATP phosphoribosyltransferase his1 - *S. pombe*
An01g14130	beta-isopropylmalate dehydrogenase A leu2A - *A. niger*
An02g00890	s.sim. phosphoserine transaminase SER1 - *S. cerevisiae*
An01g11930	s.sim. histidinol-phosphate transaminase his3 - *S. pombe*
An15g02490	sim. 3-isopropylmalate dehydrogenase leuB - *Sulfolobus* sp.
An15g02360	s.sim. acetylornithine aminotransferase arg8 - *K. marxianus*
An11g09510	s.sim. aspartate-semialdehyde dehydrogenase HOM2 - *S. cerevisiae*
An03g03010	hypothetical serine/threonine protein phosphatase – *M. aeruginosa*
An04g01610	s.sim. acetolactate synthase regulatory chain ILV6 - *S. cerevisiae*
An04g00430	s.sim. branched-chain-amino-acid transaminase BAT2 - *S. cerevisiae*
An16g07400	s.sim. methylenetetrahydrafolate dehydrogenase MTD1 - *S. cerevisiae*
An04g05490	s.sim. protein fragment SEQ ID no. 4306 patent EP1033405-A2 - *A. thaliana*

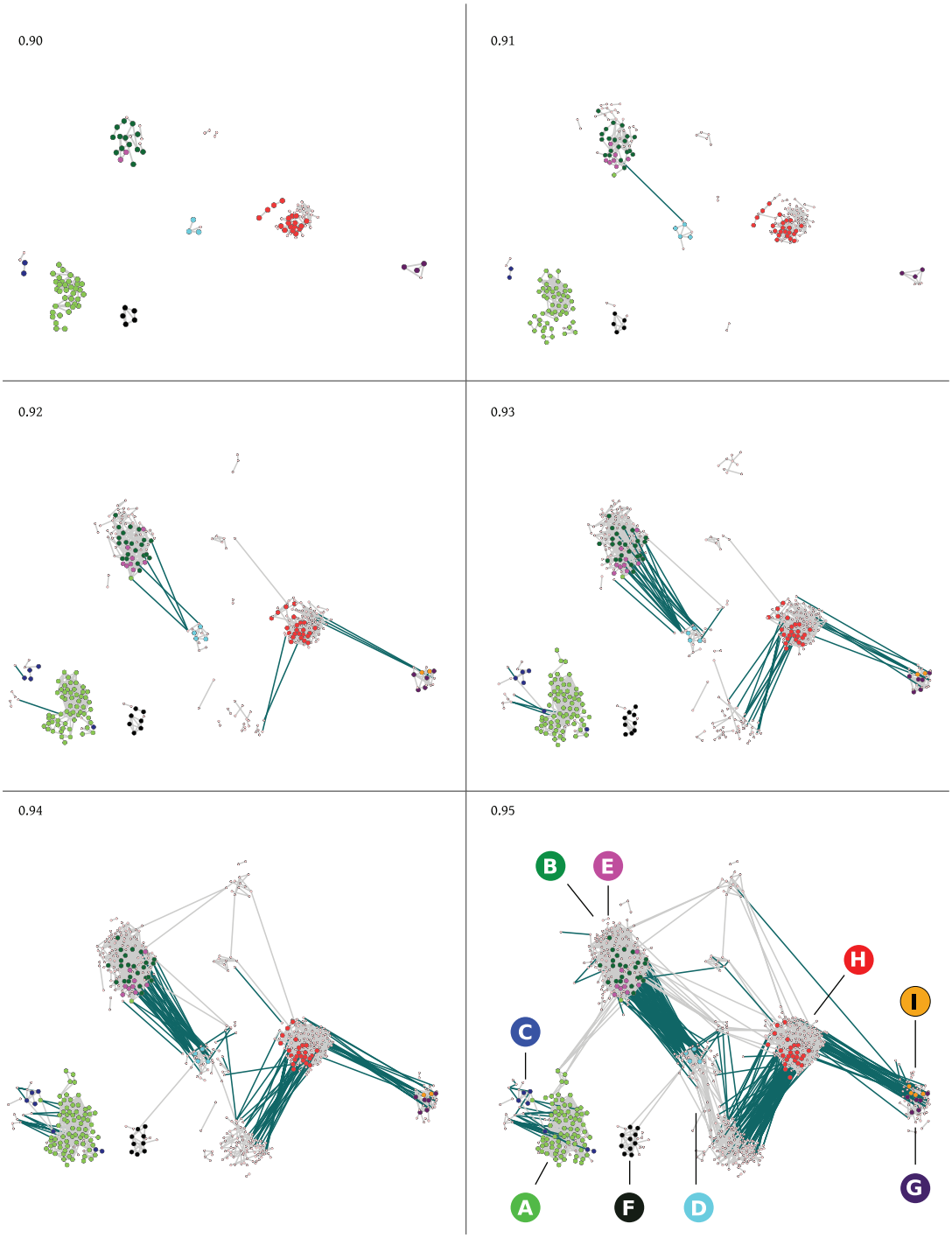


Figure 3 - The gene co-expression network of the mild perturbed conditions data set for 6 threshold settings of ρ. Genes are represented by circles, while lines represent a ρ value above the threshold. Positive ρ values are shown as solid gray lines while negative ρ values are represented as green lines. The numbers in each left corner give the ρ cut-off value used. Genes that partner in the 718 gene pairs that are present in all three networks are colored. Coloring is based on the 9 modules identified in the 718-gene pairs network: Module A, light-green; B, dark-green; C, dark-blue; D, light-blue; E, pink; F, black; G, purple; H, red; I, yellow.

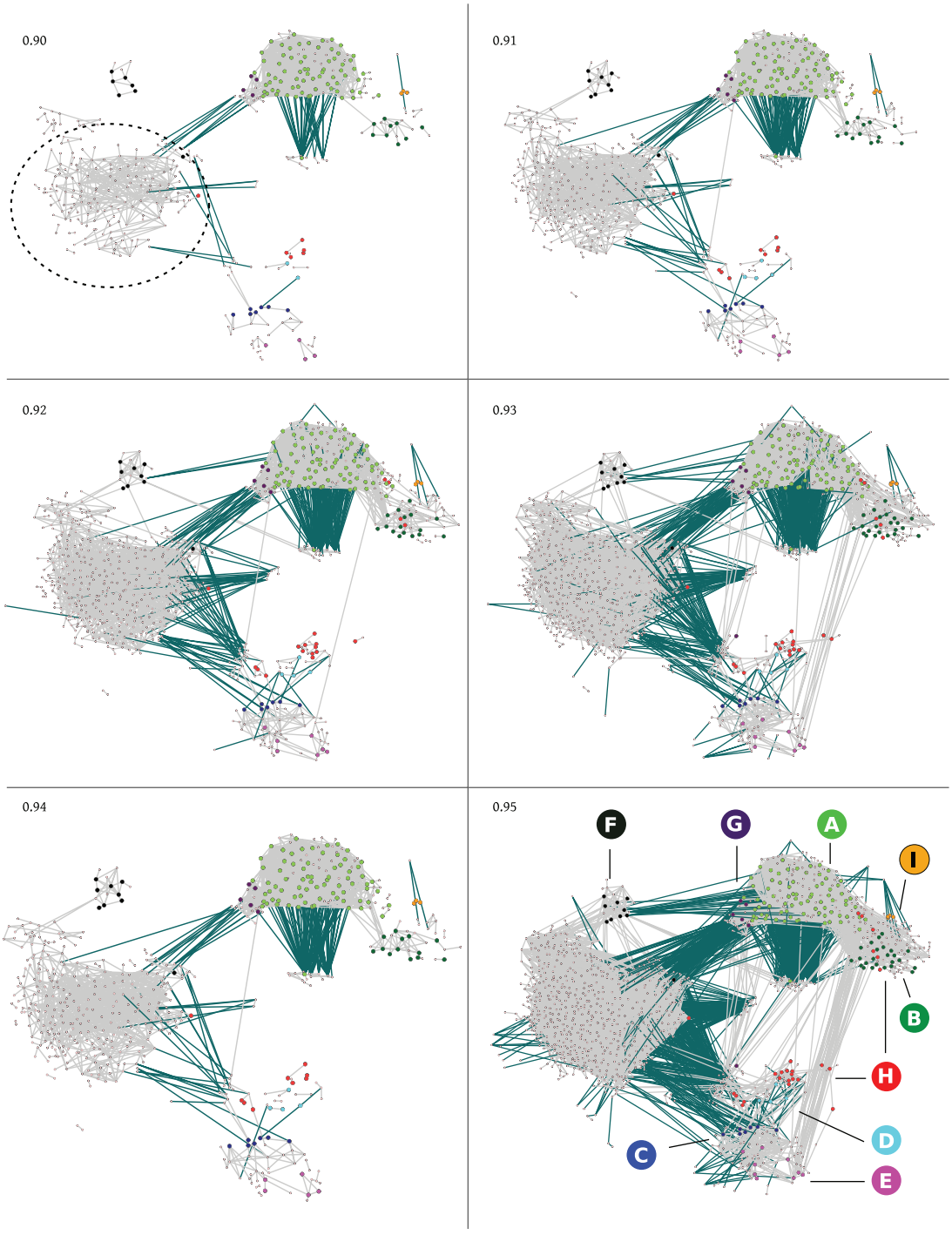


Figure 4 - The gene co-expression network of the strongly perturbed conditions data set for 6 threshold settings of ρ. The dotted circle at ρ 0.95 indicates the large group present only within the strongly perturbed conditions (see text for details). For coloring scheme, see Figure 3 legend.

average gene connectivity. Network theory stresses the relative importance of highly connected nodes [3, 6]. In *S. cerevisiae* gene co-expression networks, a positive relationship between gene connectivity and the proportion of essential genes was observed [9]. Twenty-five percent of the 2,773 genes in the used subset have a *S. cerevisiae* ortholog that is considered essential [37]. However, for the 361 genes that are present in all three networks, no over representation of *A. niger* orthologs of *S. cerevisiae* essential genes is found.

The gene co-expression networks have similar modules

We suspected that the gene co-expression networks constructed from the mildly and strongly perturbed data sets might be different given their differing cultivation conditions. This notion was supported both by the differing distributions of ρ values for each data set (Figure 1 A), and by the presence of only a minority of strongly co-expressed genes shared between the data sets (Figure 2). While 718 gene pairs are common between all three networks (Figure 2), it is well possible that their relative location to each other within each network is different given the widely varying experimental conditions used.

We therefore examined the location of these strongly correlating gene pairs within the mildly and strongly perturbed conditions networks. For this examination, an assisting co-expression network was constructed, using solely the 718 gene pairs that are common to the three networks. Nine modules of interconnected genes with similar expression profile were identified within this assisting network. These modules are labeled A through I for convenient reference, and the genes within each module were assigned an identical color. Next, gene correlation networks were drawn for both data sets (Figures 3 and 4), using a series of ρ threshold values from 0.90 to 0.95. The coloring scheme of the genes within the 9 modules was superimposed on these networks.

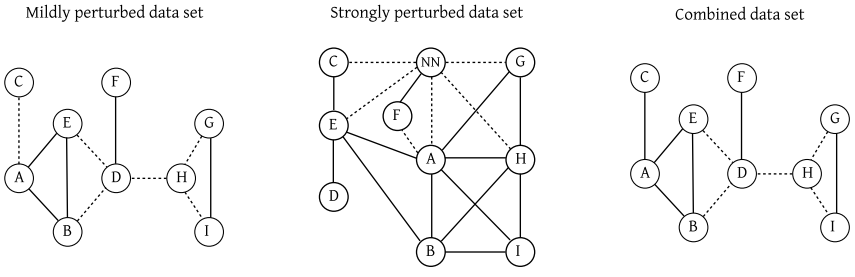


Figure 5 - A schematic representation of the relative location of the modules A to I within the gene co-expression networks at ρ 0.90. Solid line indicates that the sign of average ρ for gene pairs with members in two modules is positive, while dotted line indicates that this sign is negative. NN, the large group present only within the strongly perturbed conditions network.

The colored genes retain their local proximity: alike-colored genes are mostly grouped together both in the mildly and strongly perturbed conditions network. However, the location of the colored modules relative to each other appears not identical between the two networks (Figure 5). Genes of some modules appear to be connected by positive connections in both networks, such as the genes of modules A-B/E (modules are described in detail below). Yet other modules, such as A-C, appear connected in the mildly perturbed conditions and combined data sets networks but are not connected under strongly perturbed conditions.

The network constructed from the combined data sets (not shown) is highly similar to that of the mildly perturbed network (Figure 5). Only the A-C modules appear linked by mainly negative correlations in the mildly perturbed conditions network and by mainly positive correlations in the combined data sets network. However, the number of gene pairs that connect these modules is below 15 and contains both positive and negative correlations in both networks.

A clear difference observed between the mildly and strongly perturbed networks is the presence of a large and apparently heterogeneously structured group of gene pairs with the latter network (Figure 4, circle). Gene pairs within this group of gene pairs are loosely associated at high ρ threshold settings but become more connected at lower threshold of ρ. This module contains 600 genes that take part in a combined 5,500 correlated gene pairs. Over 90% of these correlations are between genes that are located within this module. In effect, this module contains just over half of the genes as well as over half of the correlating gene pairs connections (both 53%) of the strongly perturbed data set at a threshold ρ of 0.90.

We conclude from this visual inspection of the gene co-expression networks of the mildly and strongly perturbed conditions that these networks indeed are different. However, much of the differences are observed in the relations between modules. Within each module, its genes with correlating expression profiles are in most cases the same for all three networks. In addition, a large module is observed within the strongly perturbed network only, but this module appears to have relatively few connections with other modules.

Modules relate to biological functions

The basis of 'guilt by association' heuristics [278, 289] is that genes with similar expression level profiles often encode proteins that are involved in a similar biological process. Therefore, we looked at indications for biological processes that are overrepresented in modules that are observed in the gene co-expression networks, using

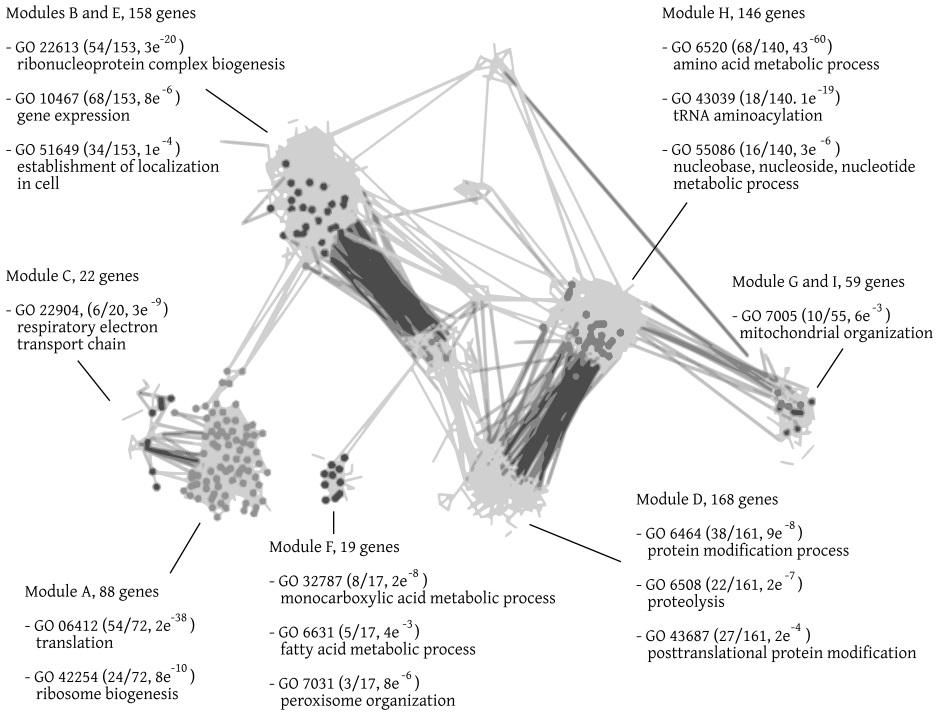


Figure 6 - Enriched biological processes for genes within the modules identified in the combined data sets network are indicated using their *S. cerevisiae* orthologs. The number of *A. niger* genes per module is indicated after the module code. The GO-number points to the observed Gene Ontology process. Between brackets, the number of genes with that annotated Gene Ontology process relative to the total number of genes queried is given, followed by a p-value that gives the likelihood that the identified GO process is found by chance. Genes that encode conserved proteins but have no *S. cerevisiae* ortholog make up the difference in the total number of *A. niger* genes per module, and the number of genes queried for Gene Ontology enrichment.

the *S. cerevisiae* Gene Ontology vocabulary. This approach applies to lists of genes instead of to individual genes. To also give an impression on the proteins encoded by the individual genes present within these modules, the annotation of genes that form the 718 gene pairs conserved in all three networks is given in Table 1.

In addition to an analysis of Gene Ontology terms, we examined the upstream regions of genes within these modules for conserved upstream elements that hint to co-regulation by a common transcription factor. For this analysis, the mildly perturbed conditions network was taken at a ρ threshold of 0.90.

Indeed, when using the Gene Ontology annotations, an overrepresentation of similar ontology terms is found for genes that are present in some modules (Figure 6). Also, conserved sequences were identified for genes of some modules.

Module H contains an overrepresentation of genes predicted to encode proteins involved in amino acid metabolic processes, including amino acid biosynthetic enzymes and tRNA-ligases. For 69 of 160 genes (50%), the conserved sequence 5'-TGA-(C/G)-TCA is identified. When a single variable nucleotide is allowed, 114 genes now contain such degenerated sequence in their upstream region (p-value 1.2 e^{-5}). The DNA-binding protein CpcA binds to this sequence [282]. This transcription factor is a regulator of cross-pathway control in *A. niger*. Upon amino acid starvation, CpcA co-ordinates a transcriptional response by derepressing transcription for many genes encoding enzymes involved in amino acid biosynthetic pathways, as well as enzymes involved in nucleotide biosynthesis.

Module F consists of genes encoding proteins involved in fatty acid metabolism or peroxisome organization. Genes encoding the peroxins Pex6, Pex10, and Pex11, as well as the bifunctional enzyme that catalyzes the second and third step of fatty acid β-oxidation FoxA, are in this module. We identified the conserved sequence 5'-CCGAGG within the upstream region of 12 of 16 genes of this module. Its reverse complement sequence has been shown to be present in a large number of genes predicted to encode proteins involved in fatty acid metabolism and peroxisome proliferation in filamentous fungi [120]. Hynes and co-workers showed that two transcription factors involved in fatty acid utilization, FarA and FarB, both bind to this sequence [120].

Module A contains mostly cytosolic ribosomal protein-encoding genes. In the upstream region of 43% of the 88 genes within this module, the conserved sequence 5'-CGACAA (p-value $4e^{-6}$) was identified, while the core sequence 5'-CGAC was found upstream 80% of the genes (p-value $2e^{-4}$). This sequence does not resemble any of the known binding sites associated with ribosomal proteins in *S. cerevisiae* or *S. pombe* [250].

The modules B and E are linked together and were analyzed as a whole. In the B/E module, genes categorized by the high-level Gene Ontology term 'gene expression' are overrepresented. For example, this module contains genes that encode putative RNA helicases, spliceosome assembly proteins, and 16 putative translation initiation factors. We identified the sequence 5'-GGCCGCG for 118 of 160 genes (p-value $8e^{-4}$). This upstream element is located 400 base pairs or more away from the gene's start site for 60% of these 118 genes. No transcription factor is known to recognize this sequence.

Next to the above-mentioned sequence motifs, we identified an overrepresentation of pyrimidine-rich sequences upstream of genes of module B/E, as well as for a large module that is solely present within the strongly perturbed conditions network. For 70% of the 600-member group of this latter module and for 80% of genes of module B/E, such CT-rich sequences are found in upstream regions. However, it should be noted that CT-rich regions are relatively common in in upstream regions of filamentous fungi and that they are mostly related to the position of the transcription start site [208].

No overrepresented biological processes were identified for genes within the large module that is solely present within the strongly perturbed conditions network. However, transcript levels for genes within this module are elevated up to 20-fold in 13 microarrays. These microarrays are derived of samples taken after depletion of the carbon source within those fermentations. In contrast, for those microarrays for which transcript levels are not elevated, a carbon source was present in the fermentations.

Discussion

Variations in the timing, location, and levels of gene transcription, mRNA translation, and protein maturation have considerable consequences for a cell. For understanding the physiology of *Aspergillus niger*, not only knowledge of individual components of the cellular system is required but also insight into the interactions and combined activity of these components. The availability of *A. niger* specific whole-genome microarrays allows for the assessment of global gene transcript levels for this filamentous fungus.

This study did not aim to query the transcriptome from the perspective of investigating a single cellular process. Rather, we sought to gain insight into the global effects that different cultivation conditions have on cellular processes.

In contrast to studies that evaluate transcription on a per-gene basis by comparing transcript levels between experimental conditions (including the studies described in Chapter 3 and 4 of this thesis), this study emphasizes the relations of a gene with other genes. Underlying our approach is the notion that genes with similar transcription profiles are thought to be more likely functionally associated. For example, such genes may encode proteins that are subunits of structural complexes, that participate in the same biochemical pathway, or that are regulated by a common mechanism. The absolute transcript levels of a gene are not taken into account in this approach. For example,

transcript levels for the 88 mostly cytosolic ribosomal protein subunits of module A differ by no more than 1.5-fold, and on average only 1.1-fold, relative to each other. Yet, a relationship between these genes is present, as their expression profiles are highly similar.

Our base questions were simple: (i) are there differences in global co-expression profile patterns when *A. niger* is cultured under distinct experimental conditions; and (ii) can these global co-expression profile patterns be used to infer underlying regulatory mechanisms of gene transcription?

The co-expression networks appear biologically plausible

Gene co-expression correlations were calculated for a subset of genes, namely those genes that encode proteins conserved between 15 or more fungi or that have a *S. cerevisiae* ortholog. Our focus on a subset consisting of conserved protein-encoding genes might affect the constructed networks. For example, their evolutionary conservation might be indicative of a relatively more essential cellular function and might be biased towards core processes that are not organism-specific, such as gene transcription or protein synthesis. This higher evolutionary conservation might also result in a network biased towards more highly connected genes. Obviously, we realize that this selection excludes an important group of fungal protein encoding genes, such as those genes that encode secreted biopolymer-degrading enzymes such as xylanases and cellulases.

We have compared two global properties of our networks with that of previously described gene co-expression networks: gene pairs distribution and modularity [17]. For the three networks described in this study, the distribution of the number of co-expressed profiles per gene can be described by a power-law distribution with a connectivity exponent γ around 1.2 (Figure 1). Similar values are found for other gene co-expression networks: a 4,077 -genes network of *S. cerevisiae* has γ around 1.0 [268], whereas this value ranged between 1.1 and 1.8 for gene co-expression networks for six distinct organisms [25]. The observed connectivity distribution does not indicate any adverse effect due to the subset used, such as for example a tendency for evolutionary conserved protein-encoding genes to have a higher than average gene connectivity.

Furthermore, the networks drawn in Figures 3 and 4 appear modular. At high ρ thresholds, few genes with a highly similar expression profile group together in distinct modules. When the ρ threshold is slightly lowered, additionally added genes have a preference for attachment to these modules rather than for initiating new modules. This modularity is also present in gene co-expression networks of other organisms, although the degree of modularity appears different per organism [25].

The modules found are enriched for biological processes – albeit that these are rather 'high-level' descriptions (*e.g.* module B/E contains genes categorized by the Gene Ontology process 'gene expression'). Interestingly, the upstream regions of genes within some modules are enriched for common conserved sequence motifs. The presence of these putative *cis*-regulatory elements suggests that shared transcriptional regulators contribute to gene co-expression. Indeed, for four modules, conserved upstream elements have been found. Two of these conserved regulatory upstream sequences have been demonstrated experimentally to be binding sites for transcription factors that relate to the biological processes enriched for in these modules: the CpcA transcription factor that relates to amino acid biosynthetic genes, and the FarA and FarB transcription factors that relate to fatty acid metabolism and peroxisome formation.

The apparent power-law distribution of co-expressed gene pairs, the observed modularity of the networks, and the confirmation that modules are enriched for genes related to biological processes by conserved regulatory elements, show that the constructed *A. niger* gene co-expression networks have biological relevance.

Similar modules, connected differently between networks

Figures 3 and 4 suggest that much of the underlying network structure is alike for both data sets. Genes that are located within a module under mildly perturbed conditions generally cluster with (many of) the same genes under strongly perturbed conditions. From a biological perspective, co-expression of genes involved in the same biological process makes sense; for example, when constructing a ribosomal complex, co-expression of all genes that encode for ribosomal sub-units is required to achieve this goal. The biological processes observed are all essential for a cell's survival: ribosomes are required for protein synthesis, genes must be transcribed, and amino acids must be synthesized *de novo* when growing on minimal medium.

But while these base modules appear largely similar for our gene co-expression networks, it is the inter-modules connections that differ (Figure 5). Negative correlating gene pairs appear over-represented for connecting the modules. For example, for 299 gene pairs of the mildly perturbed conditions network (ρ 0.90), either gene partner is located in a separate module. Of these, 240 (80%) have a negative sign of ρ. In contrast, negatively correlating gene pairs constitute 8% of all gene pairs in this network.

Our current understanding on the nature of these connections is limited and we can only speculate onto the biological principle of these relationships. The positive sign of connections between module A ('ribosome biogenesis') and B/E ('gene expression') appears to make sense as both processes are required to synthesize proteins. And also, the negative sign of correlations between modules H ('amino acid metabolic process')

and D ('proteolysis'/'post-translational protein modification') can be reasoned – but these modules appear not connected under strongly perturbed conditions. For this network analysis, only the strongest correlating gene expression profiles (ρ of 0.90 or above) resulted in constructing these networks. As the majority of correlating gene pairs are intra-modular (and are likely to be part of a similar biological process), it is well possible that many of these inter-module gene pairs are present in both data sets but are below our ρ threshold setting in one of them.

Our preliminary inspection of these genes involved in inter-module correlated gene pairs suggests that so far, no clear function could be assigned to their encoding proteins (*e.g.* 'similarity to GTPase activating protein'). Perhaps, these genes or the mechanisms that result in their actual gene transcription are involved in the 'communication' between different biological processes. We consider it worthwhile to pursue a more detailed investigation on the role of these genes within the cell, with an emphasis on their potential role as communicators for biological processes.

Conclusion

This study shows that gene co-expression networks can be constructed for *A. niger* using data of as much as 20 microarrays. Within these networks, modules are detected that consist of genes that participate in similar biological processes. Such modules remain intact under different cultivation conditions; however, the localization and connection of these modules within the networks of these conditions is different. For genes within modules, conserved regulatory sequences are found, strongly suggesting that these genes are under control of a common regulatory transcription factor; in some cases this is corroborated by experimental evidence. These findings validate the biological relevance of the networks and modules found. The conserved upstream sequences for which no function is known as yet, may serve as leads to unravel the underlying principles that provide the basis of these biological networks.

Sample name	pH	Carbon source	Nitrogen source	Nitrogen conc. [mM]	Inducer [mM]	Sampling time	Growth phase
			Strongly perturbed conditions				
4G 4NO3-1	4	glucose	$NaNO_3$	282.4	---	66	LS
4G 4NO3-2	4	glucose	$NaNO_3$	282.4	---	96	LS
4 G 8NO3	4	glucose	$NaNO_3$	564.8	---	53	LE
4 G 4NH4	4	glucose	NH_4Cl	282.4	---	57	LS
4 G 8NH4-1a (a)	4	glucose	NH_4Cl	564.8	---	36	LE
4 G 8NH4-2a (a)	4	glucose	NH_4Cl	564.8	---	36	LE
4 G 8NH4-1b (b)	4	glucose	NH_4Cl	564.8	---	60	LS
4 G 8NH4-2b (b)	4	glucose	NH_4Cl	564.8	---	60	LS
4 X 4NO3	4	xylose	$NaNO_3$	282.4	---	66	LS
4 X 8NO3	4	xylose	$NaNO_3$	564.8	---	91	LS
4 X 4NH4	4	xylose	NH_4Cl	282.4	---	60	LS
4 X 8NH4	4	xylose	NH_4Cl	564.8	---	66	LS
5 G 4NO3	5	glucose	$NaNO_3$	282.4	---	48	S
5 G 8NO3	5	glucose	$NaNO_3$	564.8	---	49	LE
5 G 4NH4	5	glucose	NH_4Cl	282.4	---	35.25	LE
5 G 8NH4	5	glucose	NH_4Cl	564.8	---	35	LE
5 X 4NO3	5	xylose	$NaNO_3$	282.4	---	93.5	S
5 X 8NO3	5	xylose	$NaNO_3$	564.8	---	112	S
5 X 4NH4	5	xylose	NH_4Cl	282.4	---	41	LE
5 X 8NH4	5	xylose	NH_4Cl	564.8	---	47.5	LE
			Mildly perturbed conditions				
29 (c)	3.5	sorbitol	$NaNO_3$	70.5	D-xylose (0.1 mM)	14	E
44 (c)	3.5	sorbitol	$NaNO_3$	70.5	D-xylose (0.1 mM)	14	E
52 (d)	3.5	sorbitol	$NaNO_3$	70.5	sorbitol (0.1 mM)	14	E
76 (c)	3.5	sorbitol	$NaNO_3$	70.5	D-xylose (0.1 mM)	14	E
86-1 (c)	3.5	sorbitol	$NaNO_3$	70.5	D-xylose (0.1 mM)	14	E
96 (d)	3.5	sorbitol	$NaNO_3$	70.5	sorbitol (0.1 mM)	14	E
Triton-0 (d)	3.5	sorbitol	$NaNO_3$	70.5	---	14	E
Triton-0.5	3.5	sorbitol	$NaNO_3$	70.5	0.002% Triton-X-100	14.5	E
Triton-1	3.5	sorbitol	$NaNO_3$	70.5	0.002% Triton-X-100	15	E
Triton-2	3.5	sorbitol	$NaNO_3$	70.5	0.002% Triton-X-100	16	E
Olive-0 (d)	3.5	sorbitol	$NaNO_3$	70.5	---	14	E
Olive-0.5	3.5	sorbitol	$NaNO_3$	70.5	olive oil (1 mM)	14.5	E
Olive-1	3.5	sorbitol	$NaNO_3$	70.5	olive oil (1 mM)	15	E
Olive-2	3.5	sorbitol	$NaNO_3$	70.5	olive oil (1 mM)	16	E
DGDG-0 (d)	3.5	sorbitol	$NaNO_3$	70.5	---	14	E
DGDG-0.5	3.5	sorbitol	$NaNO_3$	70.5	DGDG oil (1 mM)	14.5	E
DGDG-1	3.5	sorbitol	$NaNO_3$	70.5	DGDG oil (1 mM)	15	E
DGDG-2	3.5	sorbitol	$NaNO_3$	70.5	DGDG oil (1 mM)	16	E
Wheat-0 (d)	3.5	sorbitol	$NaNO_3$	70.5	---	14	E
Wheat-0.5	3.5	sorbitol	$NaNO_3$	70.5	wheat oil (1 mM)	14.5	E
Wheat-1	3.5	sorbitol	$NaNO_3$	70.5	wheat oil (1 mM)	15	E
Wheat-2	3.5	sorbitol	$NaNO_3$	70.5	wheat oil (1 mM)	16	E

Materials and Methods

Culturing

Mildly perturbed conditions: *A. niger* 872.11 (Δ*argB pyrA6 prtF28 goxC17 cspA1*) is derived from CBS 120.49. All media were based on Pontecorvo's minimal medium (pMM) [204], contained 100 mM sorbitol as carbon source with uridine and arginine as supplements. Five glass 2.5-liter fermentors (Applikon) with 2.2 liters of pMM were kept at a constant temperature of 30 ± 0.5 °C while fermentor headplates were kept at 8 °C. The medium was kept at a pH of 3.0 during cultivation. 1.0×10^6 of spores per mL were added to the fermentor. During germination, each fermentor was aerated through the headspace (50 L/h) and stirred at 300 rpm. When dissolved oxygen tension levels dropped below 60% for over 5 minutes, the stirrer speed was set to 750 rpm and aeration was switched to sparger inlet. In one experiment, fermentors were induced with either 0.1 mM sorbitol or D-xylose at T=14 hours (Chapter 3, [264]). In a second experiment, fermentors were induced with 1 mM of various oils at T=14 hours, and samples were taken before induction and up to 2 hours after induction (Table 2) (Chapter 4).

Strongly perturbed conditions: *A. niger* N402 (*cspA1*) [267] is derived from CBS 120.49. All media were based on Bennett's minimal medium (bMM) [23]. The carbon source used was either glucose (277.5 mM) or D-xylose (333.0 mM). The nitrogen source used was either sodium nitrate or ammonium chloride, at two different levels (282.4 or 564.8 mM (Braaksma *et al.*, in preparation). The different growth conditions are listed in Table 2. To prevent foaming, 1% (v/v) antifoam (Struktol J 673) was added to the medium. When necessary, additional antifoam was added during cultivation. Cultivations were carried out in 6.6-liter fermentors (New Brunswick Scientific) with 5.0 liters of bMM. For inoculation of the batch cultivations, baffled 500 mL Erlenmeyer flasks were inoculated with 1×10^6 spores per liter. Subsequently, the flasks were incubated at 30 °C in a rotary shaker at 125 rpm until approximately half of the available carbon source was consumed, which took four till seven days. Each flask contained 100 mL bMM (pH 6.5) supplemented with carbon source and nitrogen source, identical to the medium in the batch cultivations. The fermentors were inoculated with 4% (w/v) pre-culture. Each fermentor was sparged with 75 L/h of air with the stirrer speed set at 400 rpm at the start of the cultivation. When dissolved oxygen tension levels dropped below 20%, the stirrer speed was automatically increased to maintain oxygen tension at 20%. Stirrer speed was

Table 2 (opposite page) – Fermentation conditions for the samples used for DNA microarray analysis. LS, late stationary growth, over 10 hours of carbon depletion; S, stationary phase, carbon source will become depleted within 1 hour; LE, late exponential growth phase; E, exponential growth phase. Arrays indicated by (a) and (b) are technical replicates. Samples (c) and (d) are biological duplicates.

adapted until its maximum of 1000 rpm. The pH was controlled at the set value (Table 2) by automatic addition of 8 M KOH or 1.5 M H_3PO_4. Samples were taken either at a time point near carbon source depletion or approximately one day after carbon source depletion.

RNA isolation

Culture samples from the **mildly perturbed conditions** were filtrated and biomass was snap-frozen into liquid nitrogen and stored at -80 °C. Culture samples from the **strongly perturbed conditions** were immediately quenched at -45 °C in methanol as described previously [200], and centrifuged at -20 °C to remove supernatant. Biomass was frozen into liquid nitrogen and stored at -80 °C. A Trizol-chloroform extraction preceded total RNA extraction with RNeasy mini columns (Qiagen) according to the manufacturer's protocol for yeast. Concentration of total RNA was determined by spectrophotometry. RNA integrity was assessed on an Experion system (Biorad) for samples from mildly perturbed conditions, and on agarose gel, by its A260/A280 ratio, and on an Agilent 2100 Bioanalyzer for samples from strongly perturbed conditions.

Microarray processing

For mildly perturbed conditions: cDNA and cRNA synthesis and labeling, and array hybridization were performed following the Affymetrix users' manual [3] using the One-cycle Target Labeling and Control Reagents Kit to synthesize 15 µg of cRNA. **For strongly perturbed conditions**, the Bioarray High Yield RNA transcript labeling kit (Enzo) was used to synthesize at least 30 µg of cRNA. 15 µg of labeled cRNA was hybridized to custom-made *A. niger* arrays at 45 ˚C for 16 hours. Washing and staining was done using the Hybridization, Wash and Stain Kit (Affymetrix) using a GeneChip FS-450 Fluidics station and an Agilent G2500A Gene Array scanner. Scanned images were converted into .CEL files using MicroArray Suite software (Affymetrix).

Microarrays for the mildly perturbed conditions are deposited at the NCBI Gene Expression Omnibus database, accession numbers **GSE11405** and **GSE14285**. The library and platform information for this proprietory dsmM_ANIGERa_coll511030F GeneChip (Affymetrix) is deposited in this database under accession number **GPL6758**.

Data analysis

Intensity and flag values were derived using Affymetrix' MicroArray Suite Software version 5 (MAS5) with the target value set at 100 [1]. Probe sets flagged 'absent' in more than 80% of the microarrays per data set were discarded from the specific data sets to as their limited number of observations hampers the calculation of reliable correlations. Probe set values were normalized per microarray by dividing each probe set value by the mean signal over the whole microarray. Subsequently, the remaining signals flagged as

'absent' (thus not including the removed probe sets) were replaced by the mean absent call value divided by two as a uniform “lowest value in the data set”. Other values below this mean absent call divided by two were also set to this value. Genes were not filtered for a certain fold change threshold as the magnitude of fold change is not necessarily a measure for biological relevance [263]. In addition, the correlation analysis of the data has its own selection criterion, namely the ρ threshold value.

Correlation analysis is a common technique in microarray analysis [16, 71]. The correlations between genes were determined by the Spearman correlation coefficient ρ. This number ranges from 0 (no correlation) till either 1 (full positive correlation between expression levels) or -1 (full negative correlation between expression levels, *i.e.* perfect antagonists). The Spearman ρ is robust against outliers and mild non-linear behavior [297]. The p-value for the mild perturbed conditions data set, which is the smallest data set, was 4.12×10^{-6} for the lowest cut-off value for ρ at 0.90.

The influence of individual microarrays on the correlation analysis was evaluated by a “leave N samples out” validation. This validation procedure tests whether the strongest co-expressed gene pairs also remain the strongest co-expressed gene pairs in case microarray samples are removed from the complete set of microarrays. The following procedure was used: (i) random selection of two (for each single data set) or five (for the combined data sets) microarrays and removal from the data set; (ii) calculation of new correlation coefficients; (iii) continuation until all microarrays were excluded whilst ensuring that previously removed arrays were not removed again; (iv) repetition of this procedure for 20 times; (v) calculation of the mean correlation coefficient and variation per gene pair; (vi) selection of gene pairs matching the 2.5 and 97.5 percentiles of the mean correlation coefficient; (vii) comparison of the selected genes with the genes present in the correlation networks. For the mildly perturbed data set, 80% of its strongest correlating gene pairs fell within the 2.5 and 97.5 percentiles in this validation procedure. For both the strongly perturbed data set and the combined data set, all observed strongest correlating gene pairs belonged to the 5% strongest correlating gene pairs of the validation.

All calculations were performed on a Pentium 4 personal computer with 1GB internal memory using Matlab (The Mathworks), the Statistics Toolbox (The Mathworks), and home-made scripts.

Correlation networks were drawn in Cytoscape [232]. The length of the connecting lines does not represent the degree of correlation between the two connected genes. The coloring scheme was deduced from a network constructed from the 718 gene pairs common to the three data sets. This network consisted of 9 unconnected modules of

genes with similar expression profiles at ρ 0.90. These modules were labeled A through I, and genes within each module were assigned a color to assist localization of these genes within the networks.

Promoter analysis was done in GeneSpring, version 7.2 (Agilent) using the "Find potential regulatory sequences" tool. The promoter region from 10 to 700 bases upstream of each gene of a cluster was searched for oligonucleotides ranging from 5 to 10 bases, at maximum one single point discrepancy allowed, and correcting for local nucleotide density. The likelihood of random occurrence of identified sequences was compared relative to the upstream region of all 14,554 genes in the *A. niger* genome.

Gene Ontology

The genomes of the fungi *A. niger*, *A. fumigatus*, *A. nidulans*, *P. chrysogenum*, *N. crassa*, *M. grisea*, *S. nodorum*, *U. maydis*, and *T. reesei*, and the yeasts *A. gossypii*, *C. albicans*, *C. neoformans*, *D. hansenii*, *G. zeae*, *K. lactis*, *P. chrysosporium*, *S. cerevisiae*, *S. pombe*, and *Y. lipolytica* were used to construct an in-house built database of orthologous protein sequences (S. Basmagi and P. Schaap, unpublished data). Protein sequences were placed into an orthology cluster based on bi-directional first-hit BLAST alignment of protein sequences of other species. The presence of a protein in such cluster of orthologs is indicative of evolutionary conservation between species and increases the likelihood that the protein is biologically functional. Conserved proteins were defined as having an ortholog in at minimum 15 of 19 species; 2,749 genes fulfill this criterion. All 455 genes that have a *S. cerevisiae* ortholog but did not meet the criterion were added. On average, these genes have an ortholog in 11 species. Gene ontology terms, available from *S. cerevisiae* orthologous genes per cluster, were browsed at the *Saccharomyces* Genome Database website [109].

Acknowledgements

The authors would like to thank DSM Food Specialties for access to the DSM *A.niger* Affymetrix microarrays, and for providing the microarrays to MB. The Kluyver Centre for Genomics of Industrial Fermentation, which is supported by the Netherlands Genomics Initiative (NROG), is gratefully acknowledged for financial support of this research for MB and RB.

Chapter 6

A method for analysis of gene regulation using a chimeric $Zn(II)_2Cys_6$-type transcription factor

Douwe van der Veen,
Violeta Zorraquino Salvo, and Leo H. de Graaff

Abstract

Zinc binuclear proteins are transcription factors that are only found in fungi. The *Aspergillus niger* genome encodes 296 predicted zinc binuclear proteins, making up one-third of the total number of proteins with a DNA-binding domain. For less than 20 of these transcription factors a function has been determined.

A novel method is described to study which genes are under control of a given binuclear transcription factor. A chimeric transcription factor is constructed in which its DNA binding domain is replaced by the DNA binding domain of another transcription factor. Induction of this chimera causes the transcription factor to bind to sequences recognized by the chimeric DNA binding domain and to activate expression. The genes under control of the chimeric protein can then be identified by DNA microarray technology.

A chimeric transcription factor was constructed in which the six-cysteine region of the DNA binding domain of the xylanolytic transcription factor XlnR was replaced by its counterpart domain of the acetate regulatory protein FacB. Construction of this chimera proved to be difficult in *Escherichia coli*. A spontaneous point mutation resulted in a Ser-to-Pro mutation in a non-conserved region in the XlnR protein. Growth of transformants on D-xylose led to a segregating sector morphology while no morphological change was observed when grown on acetate, glucose, or sorbitol. Neither genes under control of XlnR nor of FacB were up-regulated after induction with D-xylose as was determined by qPCR.

The construction of a functional chimeric XlnR:FacB transcription factor was not successful. However, it seems that the point-mutated chimeric protein is activated upon D-xylose induction, as suggested by the segregation sectors in colonies of transformants grown on D-xylose plates. However, neither FacB- or XlnR-controlled genes are induced, suggesting that the allosteric switch has been affected in this particular design.

Background

High-throughput sequencing technologies have enabled routine access to genome data of over 2,000 organisms, including 148 complete or draft fungal genomes [186]. Technologies to measure global transcript [225] and protein [12] levels connect the genomic blueprint information to the physiological state of the organism. However, for a large proportion of proteins encoded, the physiological function is not clear [30]. For example, for the 5,795 open reading frames in the *Saccharomyces cerevisiae* genome that are annotated with gene ontology [67, 109], for 1,281 genes (22%) the biological process they are involved in is unknown as of January 2009. For 2,027 genes (35%) a molecular function is not known. In addition, many of the proteins whose biological process and molecular function are 'known' are poorly understood [115].

Much effort is put into elucidation of the unknown function of protein-encoding genes by computational or experimental-based methods, and combinations thereof (reviewed in [37, 79]). One strategy uses gene expression data to infer gene function by following the 'guilt by association' principle [71, 278]. This principle assumes that unknown genes that are co-expressed with genes of known function are likely to encode proteins involved in the same biological process. Although this method does not reveal the precise biochemical function of the unknown protein, it provides starting points for further investigation. The soundness of this approach has been demonstrated [289].

Genes are co-expressed as the organism has sensed - in one way or another - that it requires the proteins encoded by these genes for its current physiological state. Generally, co-expression involves co-regulation, where a set of genes is under control of the same transcription factor(s). Eukaryotic transcriptional regulation of protein-coding genes is a complex series of events. Transcription factors program the expression of genes by binding to specific sites in the genome and facilitate or repress transcription of genes under their control. In many cases, both activating and repressive mechanisms act simultaneously (for example, see [287]). The combined effects of all actors and RNA polymerase II -mediated transcription determine the final amount of mRNA transcripts [169].

One strategy to study genes that are under control of a transcription factor is chromatin immunoprecipitation [110, 212]. Unfortunately, ChIP-chips are available for a limited number of eukaryotic organisms only. Furthermore, transcription regulators might be bound to DNA but be inactive, or be bound while inhibiting gene expression. Finally, regulator activity is highly dependent on the environmental conditions of the cell, such as amino acid deprivation or osmotic stress [95]. When nothing is known about the transcription factor under study, this lack of knowledge can be a problem. An elegant

alternative strategy which does not require the use of ChIP-chips is based on DNA-binding protein-mediated transposon insertion [280]. In this method, a fusion protein containing the DNA-binding domain directs transposon integration at the transcription factor binding sites upon activation. Location of the transposon integration site is determined afterwards. Again, DNA binding does not necessarily translate into transcription of the controlled gene. Yet another approach to investigate genes under control of a transcription factor are knock-out studies, followed by phenotype analysis. This approach has been investigated for 33 zinc binuclear proteins of *S. cerevisiae* grown in 19 different conditions, including media of varying nitrogen and carbon source, and anaerobic conditions [4]. However, only broad biological functions could be assigned this way.

We envisioned a fast and flexible method to determine which genes are under control of a given transcription factor. Gene expression data is used to assess effects, as it is the expression of genes that is of physiological relevance to the organism. The transcription factor DNA-binding domain is replaced by the binding domain of a second transcription factor of the same family. Subsequently, the fungus is cultured and induced, and the chimeric transcription factor binds to the genome sequences that are recognized by its DNA binding domain. The culture is harvested and a global gene expression profile is measured. Genes under control of the chimeric transcription factor will be differentially expressed. A graphical representation of our method is given in Figure 1.

The zinc binuclear cluster motif ($Zn(II)_2Cys_6$) is a DNA-binding domain named after its 6 cysteine residues which bind 2 zinc atoms and co-ordinate folding of the DNA-binding

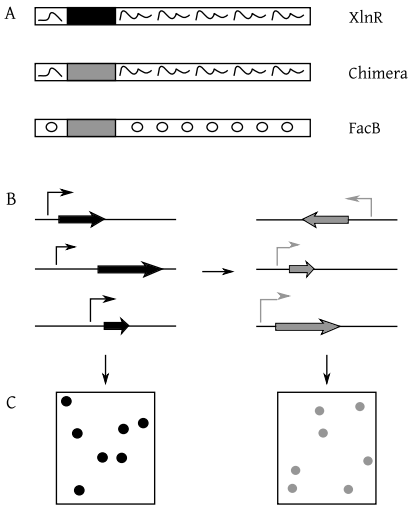


Figure 1 - Schematic representation of the proposed method. (A) The DNA-binding domain of the transcription factor XlnR (represented by red box) is replaced by the DNA-binding domain of a second transcription factor (green). (B) While the native XlnR transcription factor regulate genes under its control, the chimeric transcription factor binds to the promoter region of genes to which the chimeric DNA-binding domain binds. (C) After induction, global gene expression levels are measured by DNA microarray technology, and genes under control of the chimeric transcription factor can be identified.

domain [261]. Members of this protein class, which includes the canonical *S. cerevisiae* Gal4 protein [97], have thus far only been identified in fungi. Zinc binuclear proteins are involved in many processes including primary and secondary metabolism, drug resistance, and mitotic and meiotic development (reviewed in [164, 255]). Crystal structures of the DNA binding domain of various protein members have been solved ([164, 170]). In yeast genomes, between 54 [4] and 70 [168] putative zinc binuclear cluster transcription factors have been identified, which amounts to around 25% of the total number of transcription factors identified in these species.

The *Aspergillus niger* regulator of xylan-degrading enzymes, XlnR, is a member of the zinc binuclear cluster class of proteins [271]. The xylanolytic system is induced by xyloside moieties including D-xylose. The DNA binding sequence recognized by XlnR has been deduced from electrophoretic mobility shift assays and DNaseI footprinting [270], and genes under XlnR control have been identified [271]. This transcription factor has been functionally characterized including the determination of its cellular location [98], and various transformants including a constitutive XlnR are available [99]. The $Zn(II)_2Cys_6$ DNA-binding domain is located at the N-terminal site (amino acids 55-81), followed by a predicted nuclear localization signal and a coiled-coil domain [270]. The regulation domain of this protein is at the C-terminal site [98].

To make full use of the extensive amount of information available, we have selected XlnR as protein platform to investigate the feasibility of our method. The FacB transcription regulator was selected to donate its DNA binding domain in this proof-of-principle experiment. FacB regulates genes involved in acetamide and acetate utilization [118, 258], and this gene and expression of genes under its control have been studied in *Aspergillus nidulans* (*e.g.* [119]). DNA-binding sites have been identified for FacB [256]. Finally, the genes under control of XlnR and FacB are related to distinct physiological processes (degradation of a polysaccharide versus utilization of acetate), which reduce potential cross-over effects that might occur.

Results

Construction of the XlnR-FacB chimera transformant

To have a flexible system to create transcription factor chimera constructs, we sought to insert unique restriction enzyme sites flanking the *xlnR* DNA binding domain encoding region by site directed mutagenesis (SDM). *In vitro* SDM is a widely used technique to insert mutations at defined sites in plasmids and is used routinely in our laboratory. For the introduction of the *XhoI* site only a single G to A mutation needed to be mutated

whereas for the other restriction enzyme sites 3 base pair transitions were required. However, all attempts to introduce restriction enzyme sites at the flanks of the binding domain using SDM were unsuccessful. Variations in all parameters likely to affect SDM outcome, such as the primer pairs and PCR conditions, have been tried without success.

As the SDM method proved unsuccessful, a 500-bp region of *xlnR*, in which the DNA binding domain encoding region is replaced by that of *facB*, was *in vitro* synthesized. Propagation of the synthetic vector in *E. coli* went without problems, just as propagation of any other *xlnR*-containing plasmid. However, attempts to replace the native *xlnR* binding domain encoding region with the *xlnR:facB* encoding sequence again proved difficult. The transformation yield was very low for this construct: none or a few colonies were obtained per transformation plate, which is in contrast to over 200 colonies for other constructs made with the expression vector used. When colonies were characterized by restriction digest analysis, in most cases the empty vector was identified. The remainder of plasmids showed unexplainable and seemingly random digestion patterns, apart from one clone, pAL145-j. Unfortunately, the plasmid sequence showed a C-to-T mutation at nucleotide position 1539, which results in a proline to serine amino acid change in the encoded chimeric protein. A schematic representation of the chimeric XlnR:FacB protein is given in Figure 2. Nevertheless, this plasmid was transformed into Δ*xlnR* strain NW199 to assess its functionality. Fifty clones were isolated after transformation, and from these isolates 10 transformants were randomly selected for further analysis.

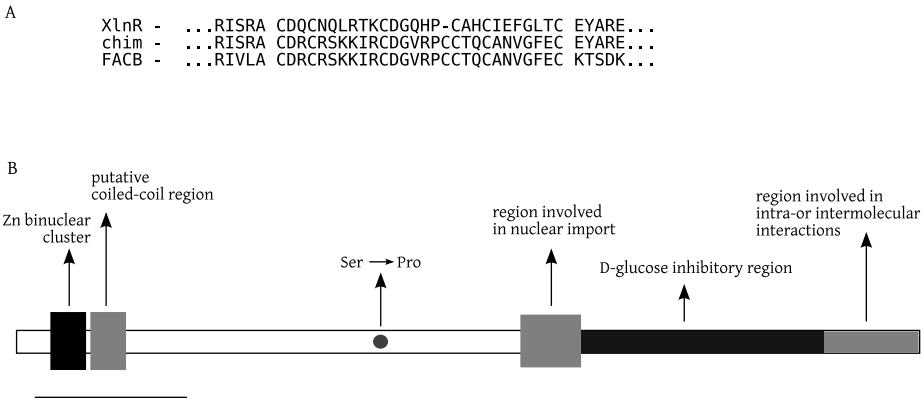


Figure 2 - XlnR transcription factor. (A) Alignment of the $Zn(II)_2Cys_6$ DNA binding domains of XlnR (amino acids 50 - 86, top) and FacB (amino acids 19-56, bottom) and the resulting chimera amino acid sequence. (B) Schematic representation of the domain organization of XlnR, adapted from [98]. The C-terminus downstream of the coiled-coil region involved in nuclear import fulfils a regulatory role, and a part of this region harbors a D-glucose inhibitory domain. The serine to proline mutation of plasmid pAL145-j is indicated as black circle.

Xylanolytic enzyme activity detection

Observation of the activity of enzymes may serve as proxy for the functionality of the transcription factor that regulates their corresponding genes. Detection of xylanolytic enzymes activity was used in a agar plate assay to test for XlnR activity. On plates that contain dyed xylan, the activity of xylanolytic enzymes releases dyed xylan oligomers. These dyed oligomers are soluble, and create by diffusion into the medium a blue halo that originates from the fungal colony [98]. Strains transformed with either the empty vector or with a plasmid containing the chimeric transcription factor were spotted onto AZCL-xylan dyed agar plates. A transformant harbouring the V756F mutation, which results in constitutive XlnR activity, and a transformant with the *xlnR* gene preceded by the upstream region of the constitutively expressed gene encoding pyruvate kinase served as controls (Figure 3).

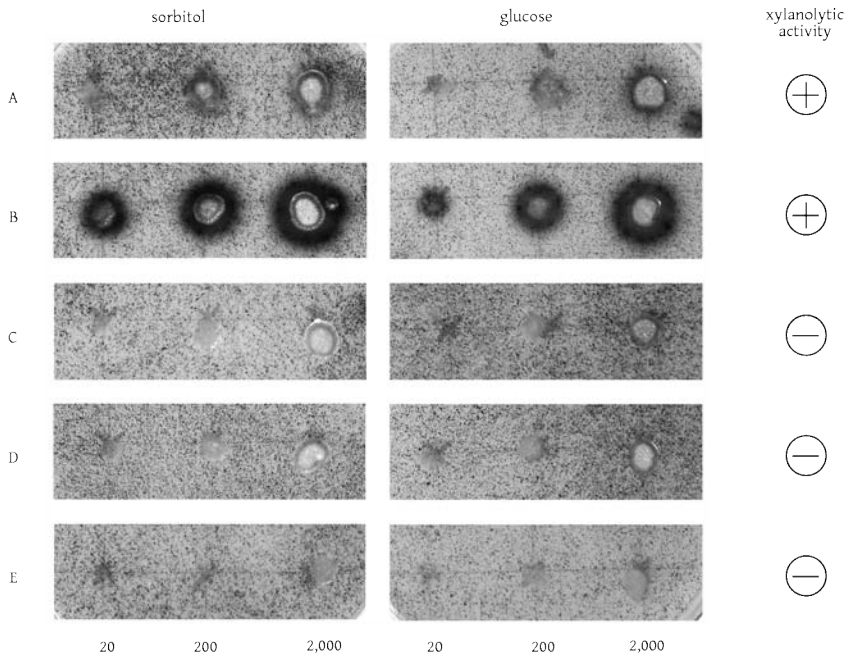


Figure 3 - Xylanolytic enzyme activity assay. A dilution series of spores (from left to right, 20, 200, and 2000 spores) was added to agar plates containing either sorbitol (left row) or D-glucose (right row) as carbon source, with a top layer of AZCL-xylan and D-xylose. Image taken after 48 hours incubation at 30 °C. Strains used: (A) V756F constitutive transformant, (B) *pki*(P):*xlnR* contruct, (C) empty vector (pAL85) transformant, (D) chimeric transformant pAL145-14, (E) chimeric transformant pAL-145-18.

Blue halos are formed both by the transformant that contains the constitutively expressed *xlnR* gene and by the transformant in which *xlnR* is placed under control of the pyruvate kinase, or *pkiA*, promoter (Figure 3, A and B). When the empty vector is transformed into strain Δ*xlnR* background strain NW199, no xylanolytic activity is detected as is expected (Figure 3, C). Two distinct colony morphologies became visible on the assay plates for transformants containing the *xlnR:facB* chimeric gene (discussed below), but for neither morphology an xylan-degrading enzymatic activity is detected. This observation is expected, as the chimeric DNA binding domain should not be able to recognize XlnR binding sites in the upstream regions of xylanolytic genes but should bind to FacB binding sites instead.

Apart from their pathway-specific regulators, both the xylanolytic and acetate utilizing regulon are under control of carbon catabolite repression [54, 134]. Both the presence either fructose or glucose have a strong repressing effect on the expression of xylanolytic enzyme-encoding genes [8]. CreA is the major protein responsible for carbon catabolite repression in *Aspergilli* [235]. The effect of CreA-mediated repression on transformants was assessed by growing the selected transformants on medium containing either sorbitol or glucose as carbon source. Sorbitol does not induce carbon catabolite repression. Indeed, when either the constitutive V756F and the *pkiA*(p):*xlnR* transformants are grown in glucose containing medium, a decreased halo size - indicating reduced enzyme activity - is observed in comparison to growth on sorbitol (Figure 3).

Induction by D-xylose alters colony morphology

Two morphologies could be distinguished within the 10 transformants that were patched onto the plates used to assay xylanolytic enzyme activity. These morphologies were present both when glucose or sorbitol was used as carbon source. Representative colonies are shown in Figure 4. For four transformants, colony morphology was not different to that of the wild-type morphology as is displayed by the control strains. Six transformants formed irregular-shaped colonies and grew slower over time. The morphological difference became profound when patches of dense spore-forming regions appeared alongside less dense regions after 7 days of incubation (Figure 4, C and D).

No such colony morphology has been observed on the transformation plates and during pure culture isolation. To examine the morphology in greater detail, spores from glucose or sorbitol containing plates were picked either from sectors of regular appearance or from sectors with sparse and irregular distribution of spores. These spores were transferred to MM plates with either sorbitol, or sorbitol and D-xylose as carbon source, and incubated for 5 days (Table 1). Of 8 sparse spore-distributed sectors that were

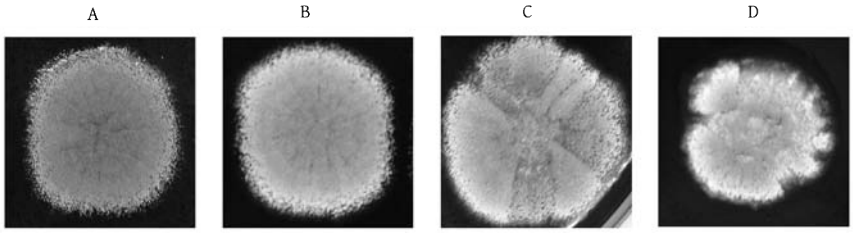


Figure 4 - Colony morphology of colonies grown on agar plates containing D-glucose as carbon source, with a lop layer of AZCL-xylan and D-xylose, after 124 hours of incubation at 30 °C. From left to right: (1) the *pki*(p):*xlnR* construct; (2) empty vector pAL85 transformant; (3) chimeric transformant pAL-145-16; (4) chimeric transformant 145-18.

Transformant	Sector type	Original carbon source	MM + xylose	MM + sorbitol
145-18	regular	glucose	no	no
145-18	**sparse**	glucose	**yes**	no
145-18	**sparse**	glucose	**yes**	no
145-18	regular	sorbitol	no	no
145-18	**sparse**	glucose	**yes**	no
145-16	**sparse**	glucose	no	no
145-16	regular	glucose	no	no
145-16	regular	glucose	no	no
145-16	**sparse**	glucose	**yes**	no
145-15	regular	glucose	no	no
145-15	**sparse**	sorbitol	**yes**	no
145-13	regular	sorbitol	no	no
145-13	**sparse**	glucose	no	no
145-10	regular	glucose	no	no
145-9	**sparse**	glucose	no	no
145-9	regular	glucose	no	no
85-2	regular	glucose	no	no

Table 1 - Assessment of the colony morphology before, and after re-plating onto fresh minimal medium plates. 'Regular': a wild-type colony morphology; 'sparse': a morphology with both dense and sparse spore regions. See text for details. The scoring in the last two columns indicate the appearance of sparse colonies after re-plating.

inoculated onto fresh plates, 5 colonies again formed sectors of regular and sparse spore distributions. However, this morphology only appeared on D-xylose containing agar plates. The 3 remaining sparse sectors as well as all spores from all regular sectors developed as regular colonies after transfer both on plates containing sorbitol, or sorbitol and D-xylose.

Transcription levels analysis

The enzyme activity plate assay indicates that XlnR functionality has not been restored in the chimeric gene transformants. To test whether the chimeric transcription factor affects FacB-controlled genes, transcript levels for genes under either FacB or XlnR control were determined by quantitative real-time PCR (qPCR). A transfer experiment was set up in which two transformants were induced with D-xylose or acetate, the respective inducers of the XlnR and FacB regulated systems. Since D-xylose affects the morphology of chimeric gene transformants on plates, both cultures derived from a 'regular' and 'sparse' colony sector were examined (Table 2).

Number	Strain	Morphology
1	NW199::*pki*(p):*gfp*:*xlnR*	regular
2	NW199::pAL85-2	regular
3	NW199::pAL145-18	regular
4	NW199::pAL145-18	**sparse**
5	NW199::pAL145-16	regular
6	NW199::pAL145-16	**sparse**

Table 2 - Transformant strains used to quantify gene expression levels.

Transcript profiles for 11 genes were determined. The *xlnD* gene encodes a β-xylosidase, and expression is under control of XlnR [271] (Figure 5, A). Expression of this gene only was elevated in the strain that contains the wild-type *xlnR* gene, when induced with D-xylose. Levels for *xlnD* were not induced in the other strains. These results are in agreement with the xylanolytic enzyme activity plate assay.

Expression of the genes encoding the two enzymes of the glyoxylate shunt, isocitrate lyase and malate synthase, is induced by acetate via FacB control in *A. nidulans* [159, 256]. Indeed, the *A. niger* ortholog of malate synthase which is encoded by An15g01860 was 20- to 40-fold up-regulated on acetate compared to non-induced conditions (Figure 5, B). Isocitrate lyase expression levels (*acuD*, An01g09270) were by 5- to 10-fold induced on

acetate. The gene encoding an carnitine/acylcarnitine carrier, *acuH* (An03g03360), is reported to be acetate-inducible in *A. nidulans* [158]. Transcript levels for its *A. niger* ortholog were elevated by 2- to 5-fold, a low but significant induction. The *A. niger* ortholog of the *amdS* gene encoding acetamidase in *A. nidulans*, An11g02980, is induced by acetate [118] and is under control of FacB [258]. Expression of this gene was not induced by acetate for all strains tested.

Expression levels for 3 endogenous reference genes, An08g05910, An14g05050, and An08g06940, were measured as control. These genes were shown previously to be stably expressed both when grown on sorbitol or when induced on a variety of carbon sources including olive oil and D-xylose (this thesis, Chapter 2). Expression of these genes did not vary upon induction (Figure 5, C). The one measurement of An08g05910 was considered a measurement outlier, given that the expression ratio for the two other other reference genes nor for any of the other genes did vary for this sample.

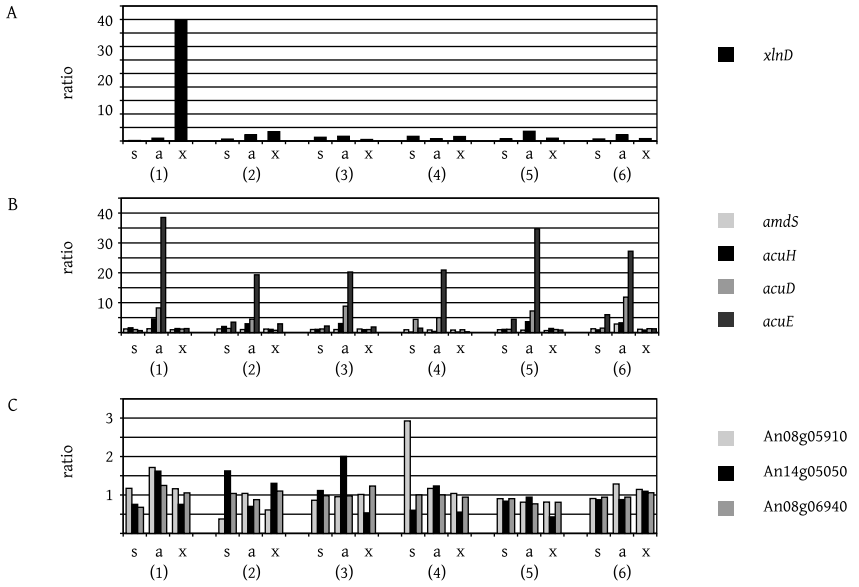


Figure 5 - Quantitative real-time PCR data for 8 selected genes. Expression is relative to non-induced condition (not induced state equals 1). Panel A: the XlnR-regulated gene, *xlnD*. Panel B: four genes regulated by FacB. Panel C: three endogenous reference genes. Conditions examined are sorbitol (s), acetate (a), and D-xylose (x). Strains used are presented in Table 3.

Discussion

Over 145 billion nucleotide bases are currently available in public databases [24], a number enlarged daily by genome and metagenome sequencing projects. With the expansion of these databases comes an increase in the number of unknown genes. For example, the metagenome sequencing project of the Sargasso Sea resulted in 1 billion base pairs, encoding over 1.2 million unknown genes [277]. Even for *E. coli*, which has been studied for decades, a physiological function for nearly 35% of its gene products is unknown [177]. Even less is known about interactions between gene products within a cell, and between those gene products and metabolites.

Co-expression of genes that encode proteins involved in similar cellular functions is a regular observation both for prokaryotic and eukaryotic organisms. Since its first observation in DNA microarray data by Eisen and co-workers [71], the 'guilt by association' principle has proven its capacity to give indications to the function of unknown proteins. Transcript profiling technologies deliver data that can be analysed for co-expressed genes. However, experimental conditions chosen by a researcher force an organism to express only genes encoding proteins that are required for that particular condition. One major drawback of this strategy is that the nature of induction for many biochemical processes is unknown. Hence, observation of co-expressed genes is one of chance alone.

In this study, we have tested a method to systematically investigate co-expression without relying on pre-determined experimental conditions. A transcription factor was modified to contain the DNA binding domain of a second transcription factor. The remainder of the protein was left unaltered, aimed to keep the modulator domain of the transcription factor functional to receive the D-xylose mediated activation signal.

Transcription factors often have a DNA-binding and modulator domain, as was shown in pioneering studies on the baker's yeast Gal4 protein (reviewed in [206]). The DNA binding domain of zinc binuclear proteins bind to a single trinucleotide sequence. These sequences appear in the promoter region either alone or as inverted or direct repeats. For example, the Gal4 binding site is 5'-CGG-N_{11}-CCG-3' [276]. The nucleotide triplet CGG is common for proteins of this family, although other triplet sequences such as GGC/A or TTA have been reported [164]. These core triplets are separated by a spacer region of varying length. The length of the spacer region is essential for proper binding of Gal4 [276]. While both Gal4 and Ppr1, an activator of pyrimidine metabolism related genes, recognize the sequence 5'-CGG-N_x-CGG-3', replacement of the $Zn(II)_2Cys_6$ domain of Gal4 by that of Ppr1 retains Gal4 binding specificity [211]. The spacer sequence also contributes to the specificity of transcription factor binding site recognition. For the two

yeast proteins Uga3 and Leu3, modification of the 4 base pair spacer sequence of Uga3 to the 4 base pair spacer sequence of Leu3 results in swapped recognition [190].

The $Zn(II)_2Cys_6$ domain of zinc binuclear proteins is followed by a coiled-coil domain, which is present in the majority of proteins of this family [164, 22]. Crystallography studies showed that the coiled-coil domain (as present in the Gal4, Put3, and Ppr1 proteins) is involved in homo-dimerization [170, 171, 247]. Absence of such coiled-coil domain results in monomer DNA binding, as is the case for the *A. nidulans* AlcR zinc binuclear protein [189]. A linker region connects both domains. DNA binding affinity experiments with Gal4, Ppr1, and Put3 chimeric proteins suggested that the linker region and the N-terminal end of the coiled coil domain are responsible for DNA binding specificity [211].

The consensus DNA binding site for XlnR is 5'-GGCTAAA-3' [270], while for FacB two consensus sequences have been identified: 5'-TCC/G-$N_{8\text{-}10}$-C/GGA-3' and 5'-GCC/A-$N_{8\text{-}10}$-T/GGC-3', respectively [256]. The XlnR protein encodes the zinc binuclear DNA-binding motif on residues 55-81, followed by residues 83-95. This sequence resembles that of a nuclear localization signal, but is not active as such [98]. For FacB, residues 24-51 encode the zinc binuclear DNA-binding motif, followed by a linker region (52-66) and a putative dimerization domain (67-101) [258].

For this study, a first approach was to swap only the six cysteine-containing DNA-binding domain of XlnR for the FacB coding sequence. The main function of this domain appears to be the recognition of the terminal triplet. A swap of only this domain would be a first test for the validity of this method, since XlnR and FacB recognize a different terminal triplet (GGC versus TCC/G or GCC/A). The FacB consensus site has rotational symmetry while the XlnR consensus site has no such symmetry. This suggests that FacB *in vivo* likely operates as a homo-dimer while XlnR functions as monomer. The absence of chimeric protein-induced transcription of FacB regulated genes after modification of its DNA-binding domain thus might be the result of an inability to recognize the FacB consensus sequence. Alternatively, the chimeric protein does recognize the FacB consensus site, but is unable to recruit the RNA polymerase transcription machinery due to a requirement for dimeric activation. The chimeric XlnR-FacB transcription factor is able to respond to D-xylose. This is evidenced by the morphological distinct phenotype seen in Figure 4. No discrimination between wild-type or chimeric transformant can be made when grown on other carbon sources than D-xylose. As only a chimera-encoding sequence that contained a point mutation was transformed, the present data does not discriminate between an effect caused by the introduced point-mutation, the DNA-binding domain swap, or by a combination of both factors.

Methods

Construction of expression vector pAL85

Plasmid pAL85 is a derivative of pRSF-1b (Novagen) which contains both bacterial and fungal selection markers and a multiple cloning site flanked by the *pkiA* promoter of *A. niger* [55] and *trpC* terminator of *A. nidulans* [180]. pAL85 was constructed as follows. An 2.7 kb *XbaI*, *BstAPI*-fragment containing the *pyrA* fungal selection marker [89] was obtained from plasmid pGW635 by PCR using primers Pyr-Fw and Pyr-Rev (Table 3). This PCR fragment was cloned into pRSF-1b yielding pAL74. Next, a 545-bp *NotI*, *Bsu36I*-fragment containing the *trpC* terminator was obtained by PCR using primers TrpC-Fw and TrpC-Rev, and cloned into pAL74. Plasmid pAL85 was obtained by cloning of a 750-bp *PfoI*, *NotI* fragment containing the *pkiA* promoter, obtained by PCR using PkiA-Fw and PkiA-Rev. Two independent pAL85 clones were sequenced with two-fold coverage for every inserted fragment.

Construction of XlnR-FacB chimera

Plasmid APL-14 ([172]) contains a full-length *xlnR* EST. A fragment containing the full-lenght coding region was amplified by PCR using primers XlnR-*SgfI*-Fw and XlnR-*DraI*-Rv (Table 3). The fragment was ligated into a pGEM-T vector (Promega) to yield plasmid pAL92. This plasmid was used as template to introduce restriction enzyme sites using the QuickChange site directed mutagenesis kit (Stratagene). Introduction of *PmlI*, *XhoI*, *BglII*, or *HindIII* restriction sites was tried using primers XlnR-*PmlI*-Fw, Xlnr-*XhoI*-Fw, XlnR-*BglII*-Fw, or XlnR-*HindIII*-Fw, and primers encoding their reverse complements. Both the manufacturer-provided protocol and variations thereof (*e.g.* variation in amount of template, primer concentration, annealing temperature), as well as the Multisite-directed mutagenesis protocol (Stratagene) were tried.

In a second strategy, the XlnR gene was obtained from *A. niger* N400 genomic DNA by PCR using XlnR-*SgfI*-Fw XlnR-*DraI*-Rv primers, and ligated into a pJET vector (Fermentas), thus yielding pAL134. A 400-bp construct in which the bases encoding the $Zn(II)_2Cys_6$ DNA binding domain of XlnR (amino acids 55-81) are replaced by those encoding the FacB DNA binding domain (amino acids 24-51) was synthesized (BaseClear BV). A 360-bp *SgfI*, *MluI* fragment was digested from this construct, and ligated into pAL134. Finally, the gene encoding the chimeric transcription factor was digested by *SgfI* and *DraI*, and ligated into overexpression vector pAL85 to yield pAL145.

Strains, growth conditions and transformation procedure

Aspergillus niger strain NW199 (*fwnA6*, *leuA5*, *goxC17*, *pyrA6*, Δ*xlnR:pIM240*) is derived from N402, a low-conidiophore mutant of N400 (CBS120.49). Strain NW199::pIM4474 harbours a functional *gfp:xlnR* fusion construct under control of the *pkiA* promoter ([99]). Strain

NW199::pIM4449-24 harbours the constitutive V756F mutant of *xlnR* [98]. In transformation experiments, strain NW199 was used as recipient strain using plasmids pAL85 and pAL145. Transformation has been carried out as described in [149], and over 50 transformants were obtained per transformation. Every transformant picked from the transformation plates was isolated in two steps to obtain a pure culture.

All media were based on minimal medium (MM) [204] and were supplemented with leucine or uridine when appropriate. MM plates containing either 50 mM sorbitol or D-glucose were poured to serve as basis layer for enzyme activity plates. For analysis of endo-xylanase activity, a top layer consisting of 0.6% (w/v) agar-agar, 0.1% (w/v) azurin-dyed and cross-linked (AZCL) xylan (Megazyme), and either 20 mM D-xylose or a mixture of 20 mM D-xylose and 6 mM D-glucose, was added to these MM plates. Blue

Primer name	Oligonucleotide sequence (5'-)	
	Construction of pAL-85	
pyr-Fw	TTTCTTTTATCTAGATTGTCGCAGG	
pyr-Rev	TCGTCTCGCAGAGTATGCTGATTGCCCCTA	
trpC-Fw	GACTGCGGCCGCTGTACAGTTT AAACAAGAATAAAACGCGTT	
trpC-Rev	CCTGAGGTGTGCATTC TGGGTAAACGA	
pkiA-Fw	TGTGGGGTCTCCGGGAATTCGTGAGAGGGT	
pkiA-Rev	CAGTGCGGCCGCGATCGCTGTACATATGGACGGATGATTGATCTCTACTGAA	
	Site directed mutagenesis primers	
XlnR-*SgfI*-Fw	ATCGGCGATCGCATGTCGCATACGAAGGATC	
XlnR-*DraI*-Rv	ATGCTTTAAACCTACAGTGCCAGCCCACTGCC	
XlnR-*PmlI*-Fw	TTCGCCGCCGAATCTCACGTGCGTGTGACCAG	
XlnR-*XhoI*-Fw	ACCTGCGAGTATGCTCGAGAACGCAAGAAGCG	
XlnR-*BglII*-Fw	GTTCGCCGCCGAAAGATCTGTGCGTGTGACCAGTGTAACC	
XlnR-*HindIII*-Fw	GACTGACCTGCGAGTATGCGAAGCTTCGCAAGAAGCG	
	Quantitative real-time PCR primers	
kana-qPCR-Fw, -Rev	AGCATTACGCTGACTTGACG	AGGTGGACCAGTTGGTGATT
An08g05910-Fw, -Rev	CCAGGATGAAGAGTGGGAGA	GCAGCTGGAGTGCTTCTTTC
An14g05050-Fw, -Rev	ACTCCAGAGGACAAGCAGGA	GCAGACGCATGCTCTCAATA
An08g06940-Fw, -Rev	ATCTTGCGTGACAACATCCA	CACCCTCAAGGAAGGTCTTG
acuH-qPCR-Fw, Rev	TCAAGCGCACTTTGACTC	GAATCCACATGGCAATACC
amdS-qPCR-Fw, -Rev	CAGCACCGCCAATTTGAC	CCCAGTATGACGGTACTTCC
faa1-qPCR-Fw, -Rev	CCTGGTGGCTATCTTTGG	GATCTGATGCCCTCGATG
facC-qPCR-Fw, Rev	GTCGAAGCATCGACAAAC	GGCAATCATCTCCATCTC
xlnD-qPCR-Fw, -Rev	TAATCTACGCCGGTGGTATC	TTCTTGAGCGAAGAGGAATC

Table 3 - Oligonucleotide primers used in this study. 'Fw' and 'Rev' indicate forward and reverse primers.

halos, formed by release of soluble dyed xylan oligomers from the insoluble substrate, indicate endo-xylanase activity. All plates were allowed to dry for two days at room temperature. Spore solutions of the transformant strains were diluted to 2, 20, or 200 spores per µL in Saline-Tween, and 5 µL of each solution was spotted onto the plates. Inoculated plates were incubated at 30 °C.

For the induction experiment, per strain, precultures were grown in three 1 Liter shake flasks containing 200 mL MM and 100 mM sorbitol at 30 °C and 160 rpm, after inoculation with 4×10^8 spores. After 18 hours, mycelium was pooled and filtrated gently over a Büchner funnel, washed twice with MM, and transferred to three 500 mL shake flasks each containing 150 mL MM, 50 mM sorbitol. The transferred cultures were allowed to grow undisturbed for 2 hours at 160 rpm. A 50 mL-sample was removed from the culture, and either sorbitol, D-xylose, or acetate to a final concentration of 33 mM was added. After 90 minutes, again a 50 mL-sample was removed from the culture. Samples were filtrated, and snap-frozen in liquid N_2.

RNA isolation and quantitative real-time PCR

Total RNA was isolated from ground mycelium using TRIzol Reagent (Invitrogen) followed by chloroform extraction. Isolated total RNA was dissolved in 10 mM Tris-HCl pH 8.0, heated for 2 min at 70 °C, and diluted in two steps to a concentration of 200 ± 5 ng × μL^{-1}. 1.00 µg of total RNA was spiked with 0.1 ng of a synthetic RNA transcript encoding kanamycin synthetase (Promega) followed by DNaseI-treatment. cDNA was synthesized using the Omniscript reverse transcriptase kit (Qiagen), using 4 units of reverse transcriptase enzyme in a final volume of 20 µL. cDNA synthesis reaction was at 37 °C and subsequently water was added to a final volume of 400 µL. Diluted cDNA was stored at -20 °C until use.

Oligonucleotide primers were designed using Primer3 software [218] and are listed in Table 3. For qPCR measurements, 25% (v/v) of diluted cDNA was added to 2× SYBR PCR mastermix (Abgene) and PCR primers using a CAS-1200 pipetting robot (Corbett Life Science). Final concentration for each primer was 1.4 µM. A Rotorgene-3000 qPCR machine (Corbett Life Science) was used to run the following program: 15 min at 95 °C; and 40 cycles of 15 seconds at 95 °C, 15 seconds at 58 °C and 20 seconds at 72 °C. A melting procedure ended the procedure. Determination of Ct-values and amplification efficiencies was done with the 'comparative quantitation' method of the Rotorgene software (version 6.0.19).

Acknowledgements

The authors would like to acknowledge the help of M. Ponsie, who helped with construction of the pAL85 over-expression plasmid.

Chapter 7

General discussion and concluding remarks

Douwe van der Veen

General discussion

Analysis and re-analysis of DNA microarray data

Publication and exchange of data is an integral part of science. In this manner the obtained results can be integrated with existing and future data and will contribute to science as a whole. The ability to link experimental data with information present in a myriad of databases is facilitated by advances in computer technology and informatics. Bioinformatics tools, such as systematic vocabularies like Gene Ontology [251], allow the comparison of information between organisms. Understanding an organism's complexity would be harder without this ability to compare data of various sources. For example, for *Aspergillus niger*, a mere 147 enzymatic activities have been characterized biochemically - a third of which are polysaccharide-degrading enzyme-related (BRENDA enzyme database [45], as of February, 2009). However, a recently published metabolic model of *A. niger* comprises 1190 biochemically unique reactions [10]. This is one of many examples that indicate the benefits to using data obtained 'via linking and connecting' in addition to 'organism-specific' data.

Fast-paced technological and methodological developments to investigate classes of biomolecules (often referred to as -omics technologies), and the publication of many data derived from such technologies, further stimulate this interconnectivity. In contrast with more qualitative data generated previously, the more quantitative nature of data generated by -omics technologies [135] allows for facilitated access of the latter by informatics tools. For example, Northern blotting [8] is an often-used method to investigate gene transcript levels under various experimental conditions. The resulting data are images, from which transcript level intensities are relatively difficult to quantify. Current mRNA analysis technologies including qPCR and DNA microarrays also rely on the interpretation of a visual signal, namely the fluorescence emitted from reporter molecules. However, their much greater sensitivity and the high technical reproducibility of these technologies in comparison to Northern blots allows for more reliable quantification of the output data. For example, for each probe set that is present on the *A. niger* Affymetrix microarrays that were used in this study, its fluorescent intensity signal is converted to a value ranging from zero to approximately 25,000 arbitrary units.

Whole-genome DNA microarray experiments can measure most of an organism's transcript levels in selected conditions with great sensitivity and accuracy. Such experiments are costly: measurement of a single sample using Affymetrix microarray technology is in the order of a thousand euros due to the reagents and equipment used; other types of DNA microarrays are generally less expensive. Furthermore, the time investment in the actual execution of the experiment and in subsequent data analysis is

substantial. In addition to these costs, the resulting data are more complex and less straightforward to interpret when compared to for example genome sequences. To assess the reliability of microarray data and to facilitate its re-analysis, microarray data repositories allow access to data acquired in such experiments.

Ten research articles describe the use of original *A. niger* microarray data - excluding the work presented in this thesis (Table 1). In a further two research articles, already published microarray data have been re-analyzed for alternative purposes. In the 5th study of Table 1, a mitotic recombination hotspot on chromosome III of the *A. niger* genome has been identified using classical genetics methods [262]. The analysis of transcript levels of genes within this predicted hotspot by use of microarray data obtained from glucose-grown fed-batch cultures of *A. niger* showed a high basal transcription of this region. This high basal expression indicated an open chromatin structure, which might provide an explanation for the observed high frequency of mitotic recombination for this hotspot region. In the 11th study, colony development of the filamentous fungus *Neurospora crassa* was investigated using whole-genome DNA microarrays [133]. The observed transcript levels were compared to gene expression patterns that were obtained in a study of *A. niger* colony development [153]. This cross-species examination of gene expression patterns in the colonies indicated that there are some evolutionary conserved patterns of gene expression during the development of fungal colonies.

Re-analysis of microarray data is done also in this thesis. In **Chapter 5**, microarray data originally generated to investigate experimental variation (as described in **Chapter 3**) or to describe the fungal response towards the presence of oils (as described in **Chapter 4**) were combined with microarray data aimed to better understand the proteolytic system of *A. niger* (Table 1, 16th study). These data sets were examined using co-expression analysis, an approach that indicated the presence of functional modules of co-expressed genes and suggested common regulatory mechanisms for genes in some modules. In **Chapter 4**, the expression profiles for 111 orthologs of *S. cerevisiae* genes that were identified in a transcriptome profiling study on peroxisome assembly and function were examined in *A. niger*. The two studies mentioned in Table 1 as well as the work described in Chapters 4 and 5 of this thesis show the potential of using previously generated microarray data in more than one research project. Re-analysis or re-interpretation of such data contributed to an increased understanding of all four systems under study.

Apart from obvious advantages of using microarray data in several studies, there may be disadvantages as well. Quantitative microarray data sets are of varying quality due to differences in the experimental procedures followed, the kind of experimental disturbance imposed, and the like. In addition, the generated data are more complex: in

	Experiment summary	Arrays	Cultivation	Access code	Reference
1	A subtraction library was used to examine the unfolded protein stress response to dithiothreitol (DTT)	8	S	not deposited	[163]
2	An inventory for carbohydrate degrading enzymes was made using bioinformatics. Expression levels of identified genes were examined by *A. niger* cultivation on a variety of carbohydrates.	29	S	E-MEXP-1626	[173]
3	A glucose fed-batch culture of strain CBS 513.88 was sampled at day 3 and 5, and expression analyzed.	2	F	not deposited[‡]	[196]
4	Protein secretion stress was induced by chemical treatment by DTT or tunicamycin, or by expression of a human protein, and the transcriptional and translational response (polysome analysis) was assessed.	12 + 8[†]	S, F	not deposited[&]	[91]
5	A mitotic recombination hotspot on *A. niger* chromosome III is characterized using microarray data.	R-4			[262]
6	Cell cultures were exposed to sublethal levels of the antifungal compounds caspofungin and fenpropimorph to assess the transcriptional response towards cell wall damage.	6	F	not deposited[‡]	[176]
7	The spatial gene expression within a colony was analyzed by examination of samples taken from 5 zones between the periphery and center of colonies grown on maltose or xylose.	28	S, ST	not deposited[#]	[153]
8	A DNA microarray containing *A. niger* BO-1, *A. nidulans* A4, and *A. oryzae* A1560 is developed and tested. All fungi were grown in glucose or D-xylose containing medium, and their transcriptomes assessed. A trispecies comparison reveals a conserved core of XlnR-regulated genes.	6[Ø]	F	GSE9276	[11]
9	An inventory for starch degrading enzymes was made using bioinformatics. Expression levels of identified genes were examined by cultivation on maltose and D-xylose.	12	S	E-TABM-324	[295]
10	A transcription factor regulating inulolytic genes, InuR, was identified. Microarrays were used to identify additional genes with a possible role in inulin or sucrose metabolism.	6	S	E-TABM-319	[294]
11	Colony development of the filamentous fungus *Neurospora crassa* was examined by transcript profiling and compared to similar-grown *A. niger*. Shared patterns in spatiotemporal regulation were observed.	R-7			[133]
12	Chemostat-cultured cells were grown on xylose and maltose. Transcriptome comparison reveals coordinated regulation of the secretory pathway.	12	F	not deposited[‡]	[130]
13	Cell cultures were transferred to two types of oil and to fructose, and the fungal transcriptional response towards these substrates in time was assessed.	21	S	not deposited	Chapter 2
14	The variation that occurs between fermentor cultures of *A. niger* was examined by variance components analysis of qPCR obtained data. Validation of assessment by transcript assessment of selected samples.	6	F	GSE11405[%]	Chapter 3
15	The response of fermentor-grown cell cultures towards three types of oil as well as to a non-induced control, was examined in time using microarrays.	16	F	GSE14285	Chapter 4
16	For describing regulation of the proteolytic system, fermentor-grown cultures differing in medium composition were sampled at various time points, and transcriptome and metabolome were examined.	20	F	not deposited	Chapter 5[$]

comparison to the single genome sequence of an organism there is an infinite number of transcriptomes: one for each particular cell in each particular strain for each particular environmental condition and time point. Thus, sometimes, interpretation is only meaningful in the context of the conditions under which they were generated.

To ensure that the costly investment for a microarray experiment results in high-quality data, robust experimental design and the minimization of non-experimentally induced variation is required. Such high-quality data not only aid in the specific research for which the experiment was originally designed, but also facilitate its re-interpretation in other studies.

Variation affects transcription analysis

A variety of lipids can be present in *A. niger*'s natural environment. Triglyceride oils store energy for later use in plant seeds; wax layers protect the aerial parts of epidermal cells of leaves, fruits, or non-woody stems of higher plants; and phospholipids and galactolipids are components of the membranes that surround plant cells and their organelles. Analysis of the genome sequence of *A. niger* [196] suggested that this fungus has a fully functional toolbox available for degradation and metabolization of lipids. At the start of this project, it was hypothesized that the fungus shows a distinct response towards different lipid types by activation of specific enzyme subsets. Initial growth experiments, in which *A. niger* was cultured on various lipids as sole carbon and energy source, seemed to confirm this hypothesis. Several physiological parameters, including the amount of biomass formed, the organic acids produced, the transcript levels of fatty acid metabolism related genes (Table 2), and the spectra of extracellular proteins, indeed suggested such differential response.

In addition, in the presence of both olive oil and glucose, lipolytic enzyme activity was detected in strain NW283, a strain defective in its global carbon catabolite repression system. No lipolytic enzyme activity was observed in this condition for strains with a functional carbon catabolite repression system. This finding, in combination with the

Table 1 (opposite page) - Published studies for which DNA microarray data was used were retrieved from PubMed. Array type Trispecies GeneChip "520520F", GEO platform GPL5975, is used in study 8, while GeneChip "dsmM_ANIGERa_coll511030F", GEO platform GPL6758, is used for the remaining studies. Re-use of elsewhere published microarray data is indicated by R, *i.e.* R-4 indicates that data described in experiment 4 was re-used. † 8 arrays for translational analysis. ø Only the *A. niger* arrays are listed. ‡ Supplemental data contains a list of MAS5.0 signal values for all probe sets. & Supplemental data contains lists of over-expressed genes with relative fold-change ratio. # MAS5.0 signal values for all probe sets can be found as supplementary data at http://www.bio.uu.nl/microbiology/fung/PhD-theses/AMLevin/ % Public release: September, 2009. $ M. Braaksma *et al.*, unpublished data that was used for the analysis described in Chapter 4.

observation that transcript levels of fatty acid metabolism related genes are differentially expressed, suggested control at the transcriptional level. The availability of *A. niger* DNA microarrays provided an excellent tool to assess the transcriptional response towards different lipids.

A transfer experiment was set up to compare gene expression levels between fructose, olive oil, and a galactolipids-enriched wheat oil fraction. This transfer experiment is described in **Chapter 2**, where its microarray data was used to validate candidate stably expressed genes. In short, cultures were transferred to shake flasks containing liquid minimal medium with either of the three substrates as sole carbon course, and sampled at 4, 6, or 8 hours after transfer. In addition, a non-induced control cell culture that contained no carbon source was included and sampled at 4 hours after transfer. In this concluding chapter, emphasis is given to two major factors that contribute to non-experimental variation in this data set: the emulsifier used, and the shake flask-based culturing method.

Substrate	Biomass [mg]		Oxalate [mM]			Isocitrate lyase [-]		
	24 h	72 h	3 h	24 h	72 h	4 h	6 h	8 h
no carbon source	73	27	0.1	0.4	0.5	nd	nd	nd
fructose	76	41	0.3	0.4	0.4	1	1	1
wheat oil	207	232	0.2	0.6	1.8	5	9	3
wheat oil, MGDG-enriched	334	294	0.2	0.6	8.7	24	15	18
wheat oil, DGDG-enriched	319	334	0.3	0.4	10.2	16	11	12
wheat oil, phospholipids-enriched	347	314	0.4	0.3	3.6	18	11	13
olive oil	150	328	0.2	0.3	5.3	5	9	3

Table 2 - Selected physiological parameters obtained from growth experiments where *A. niger* strain NW219 is transferred to minimal medium containing 0.5% (w/v) of various oils, 27.5 mM fructose, or no carbon source. Oxalate, the concentration of oxalate measured in the culture supernatant. Isocitrate lyase (ICL), the gene expression level of ICL relative to the expression level at the same time point of the fructose-grown culture. Data are the average of measurements of two biological samples. Wheat oil and fractions thereof were extracted from wheat gluten following [5]; see also Chapter 3) but with replacement of benzene by hexane. The MGDG-enriched oil fraction is obtained after the hexane:isopropanol 90:10 extraction; the DGDG-enriched fraction after acetone extraction; and the phospholipids-enriched fraction after methanol:water 97:3 extraction of the wheat gluten oil. Enrichment of lipids in the oil fractions was confirmed by thin layer chromatography following [4].

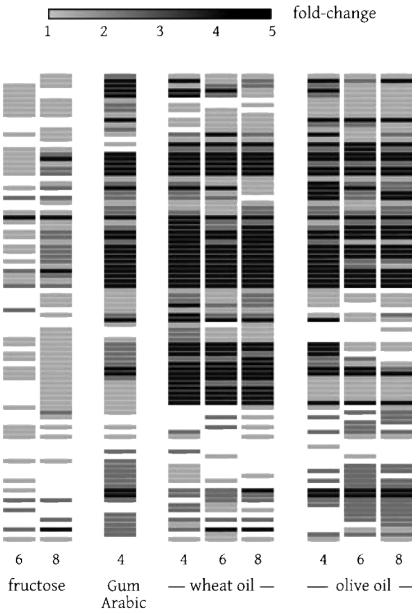


Figure 1 – Gene tree created from the expression levels of 96 genes involved in lipid degradation and metabolism, including all genes listed in Table 1 (lipases), Table 2 (peroxins), and Table 3 (β-oxidation metabolic enzymes) of Chapter 4. Expression levels are relative to the fructose-grown culture harvested at 4 hours (set at 1). Each condition is the average of two independent biological samples. Experimental details are described in Chapter 2.

Lipids are molecules with long hydrophobic tails, and an emulsifier must be used when liquid culture media are used. The emulsifier chosen for this experiment is Gum Arabic, a complex composite of polysaccharides and glycoproteins [88]. This emulsifier was added (1% w/v) to each shake flask's medium prior to transfer of the cell cultures. After the transfer experiment, microarray data analysis showed a major difference in transcriptional response between on the one hand fructose-grown cultures and on the other hand both oil-grown cultures. Of the 6,600 genes that are flagged 'present' in this experiment, one-third is over 2-fold differentially expressed on either or both oils when compared to the fructose-induced cultures. However, these genes are in overwhelming majority also differentially expressed in the control samples transferred to Gum Arabic-only medium. Figure 1 shows the expression profiles of in **Chapter 4** described genes that are involved in the four phases in fatty acid metabolism. The majority of genes that are differentially expressed on either wheat oil or olive oil, are differentially expressed also in the Gum Arabic-only samples. In addition, in the fructose-grown samples, gene expression profiles shift towards those observed for the non-induced control. Most likely, this is the result of fructose depletion followed by the consumption of the Gum Arabic. The major effect on global gene expression by Gum Arabic overshadowed the lipid-induced effects and prevented all efforts to describe the response of *A. niger* towards the added oils.

In addition to the undesired effect of the emulsifier on global gene expression, comparison of microarray samples indicated that reproducibility of biological samples was poor. Figure 2 visualizes the biological reproducibility of three pairs of microarrays. The left and middle scatter plot show microarray data from shake flask-grown cultures, while the right scatter plot shows data from fermentor-grown cultures. Between the duplicate olive oil-grown samples, 1,296 genes, or 20% of the 6,600 genes that were flagged 'present' in this study, have a signal intensity value that differs over two times between the samples. For the fructose-grown duplicate samples, 389 genes (6%) have an over two times differing signal intensity.

The large differences in global transcription levels between biological replicates of shake flask grown fungal cultures made the improvement of the reproducibility of fungal cell cultures inevitable. Furthermore, the not anticipated impact of Gum Arabic on global transcript levels pointed out the need for an experimental design that is able to absorb and account for such undesired disturbances.

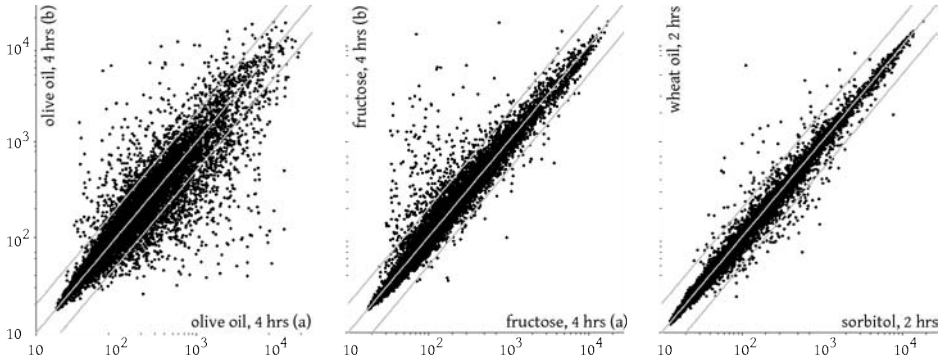


Figure 2 – Reproducibility of biological samples. The RMA normalized data of all 14,554 probe sets are plotted in a scatter plot. Signal intensity values are in arbitrary units. Left, biological replicates of cell cultures transferred to olive-oil containing medium and harvested after 4 hours. Middle, biological replicates of cell cultures transferred to fructose-containing medium and harvested after 4 hours. Both fructose and olive oil-grown cells are cultured in shake flasks; for experimental description see Chapter 2. Right, two fermentor-cultivated samples taken from different experimental condition and sampled after 2 hours: a non-induced sorbitol-grown culture is compared to a culture pulsed with wheat oil. For experimental description see Chapter 4. A gray line indicates 2-fold difference between samples.

An infrastructure to generate high-quality -omics data

To reduce non-experimental variation in future experiments, an experimental 'infrastructure' was devised (Figure 3). This 'infrastructure' is nothing more than an implementation of the familiar iterative cycle of linking knowledge to ideas: background knowledge is used to construct a hypothesis, which is tested experimentally to produce observations, which in turn add to re-evaluation of the hypothesis, after which the cycle is repeated [135, 210]. This experimental infrastructure is tailored to understanding the complexity of *A. niger*. As this complexity is approached from a experimental biology point of view in the work described in this thesis, emphasis will be placed on the more experimental biology-related left part of Figure 3.

At the center of investigation is the organism, which is *A. niger* strain 872.11 for the work described in **Chapters 3 to 6** of this thesis. This strain, the genotype of which is Δ*argB pyrA6 prtF28 goxC17 cspA1*, has some favorable characteristics over wild-type strains. Three selection markers make this strain excellent genetically amendable. The *argB* gene deletion and the mutation in the *pyrA* locus are often-used fungal transformation selection markers that confer to arginine and uridine auxotrophy, respectively. In

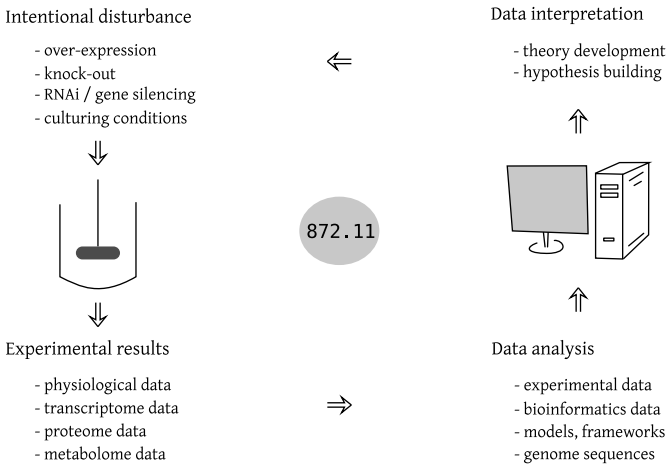


Figure 3 – Schematic representation of the experimental infrastructure envisioned in this thesis. Understanding the functioning of *A. niger* is at the center of this scheme, and is depicted here by strain 872.11 (see text for details). To design experiments aimed to understand *A. niger*, intentional experimental disturbances can be designed (top-left), of which the effects are examined using a variety of technologies (lower-left). The resulting data is analyzed to extract information, and can be connected to other data sources (lower-right). The data analysis facilitates hypothesis forming and aids in theories on the functioning of *A. niger* (top-right), which in turn can be evaluated in a new round of experiments.

addition, complementation of the *prtF28* mutation, which renders the oxaloacetate hydrolase encoding gene dysfunctional, can be used as third selection marker. In this case, selection is based on the restored ability of transformants to produce oxalic acid, which can be screened for by pH-indicator containing agar plates. In addition to these selection markers, the combined *prtF* and *goxC* mutations lead to cell cultures with greatly decreased organic acids production. This decreased acid production in turn results in lower protease activity, which is beneficial for experiments in which extracellular (heterologeous) proteins are produced. Lastly, the *cspA1* mutation results in low-conidiophore formation which improves laboratory handling and reduces the risk of strain contamination.

A variety of genetic tools is compatible with this host strain. Examples of these tools are the over-expression vector described in **Chapter 6**, which contains the *pyrA* selection marker; a knock-out system based on *prtF28* complementation that has been made during this thesis; and RNAi-based gene silencing vectors also based on the *pyrA* auxotrophic marker [193]. In addition, the majority of previously constructed vectors can be transformed into this host strain, thus connecting past and current research projects. These tools, in combination with a variety of culturing conditions, provide an excellent basis on which experimental work can be build.

The switch of cultivating *A. niger* cells in batch-fermentors instead of shake flasks resulted in the much higher reproducibility that is required for microarray studies. **Chapter 3** describes a method for the reproducible batch-fermentation culturing of *A. niger* cells. For this experimental setting, the magnitude of non-experimental introduced variation that is introduced in each individual processing step has been described. This information is important for experimental design. For example, the number of biological or technical replicates that are required to obtain a specified level of detail can be determined using these data. The effects of the improved reproducibility on microarray data can be seen in the right scatter plot of Figure 2. Microarray data derived of two fermentor-grown cultures are plotted, with 131 genes that have an over two times differing signal intensity. As opposed to the shake flask microarray samples of Figure 2, the here depicted fermentor-derived samples are not biological replicates. Instead, samples from the two most distinct fermentation as described in Chapter 4 are used: the wheat-extracted oil and the non-induced sorbitol-grown culture.

The results presented in **Chapters 3 and 4** also point to two other advantages of this experimental set-up. Fermentor-based cultivation allows for sampling the cell culture multiple times, which enables the investigation of the temporal dynamics of the culture. The different expression profiles in time of cell cultures induced with olive oil or wheat-extracted oil as described in Chapter 4, illustrate this advantage. Furthermore, this

experimental set-up enables the confrontation of fungal cells with low amounts of inducer compounds. In Chapter 3, the global transcriptional response of *A. niger* towards a pulse of only 0.1 mM D-xylose could be described. In Chapter 4, induction of batch-fermentor grown cell cultures by a 1 mM pulse of three oils resulted in differential expression of lipid metabolism related genes.

The optimization of batch-fermentor cultivation of *A. niger* yields high-quality, low-noise data. The thus obtained experimental data are combined with prior knowledge, analyzed, and interpreted (Figure 3). Various tools, including bioinformatics and statistical analyses, may assist a biologist here. In **Chapter 5**, biological information available for *S. cerevisiae* proteins is transferred to their *A. niger* orthologs by use of a database of orthologous genes. Furthermore, the correlation of expression profiles between pairs of genes is calculated in this chapter. This correlation analysis results in the detection of modules of genes. The *S. cerevisiae* data attached to their *A. niger* orthologs aided to derive putative biological functions to these correlation modules.

The final step in the iterative cycle in Figure 3 is related to hypothesis building and theory formation. In this thesis, in **Chapter 6**, it is hypothesized that the DNA binding domain and activating domain of the zinc binuclear transcription factor XlnR can function separately. A method proposed in this chapter envisions that replacement of its DNA binding domain by the DNA binding domain of other transcription factor of the same protein family, followed by D-xylose induced activation of this chimeric transcription factor, leads to transcription of genes under control of this second transcription factor. The first experimental steps to testing this hypothesis are also described in this chapter.

Concluding remarks and outlook

A living organism is the result of a complex and dynamic interplay of many molecules. Knowledge of the role and function of these individual molecules within the organism is necessary for understanding its physiology. However, also an understanding of their interactions and connections in time and place is required. The experimental biology described in this thesis adds to the unraveling and understanding of *Aspergillus niger*. The reproducible culturing of this filamentous fungus, as well as the optimization of transcription profiling technologies and methods, have enabled the production of high-quality, noise-free, quantitative data. These data have provided insight into the response of this fungus towards the sugar D-xylose and towards various oils.

The experimental infrastructure advocated in this thesis allows for the generation of quantitative data that facilitates the construction and maintenance of models that aim to describe or understand this fungus. Up till now, no models that describe the full

complexity of *A. niger* nor theoretical frameworks that are able to explain how emergent properties arise from individual biomolecular parts have been published. In describing the interactions and connections between these parts, mere intuitive and schematic representations of interactions do not suffice (an example of such schematic representation is in Figure 2, Chapter 1, of this thesis). At present, only mathematical models have the power to capture the dynamic behavior of large sets of interacting components. Quantitative data, such as generated by -omics technologies, are needed both to create models and to determine whether these models actually describe real biology.

Obviously, such models are a useful addition to the toolbox of biologists within the iterative cycle of knowledge generation. An illustrative example is described for *S. cerevisiae*, where Chen and co-workers constructed a mathematical model to describe the dynamics of cyclins within the yeast's cell cycle events [46]. Cyclins are regulatory subunits of cyclin-dependent kinases that control progression of the cell cycle in a concentration-dependent manner [81]. Predictions from this model were experimentally validated [52]. This study showed that some predictions by the model regarding cyclin abundance in various experimental situations was accurate, but also that entirely inaccurate predictions concerning some protein-protein interactions were made by this model [52]. The experimental data obtained in this study contributed to a refinement of the mathematical model, a refinement that in turn will demand its own experimental verification.

This emphasis on the quantitative nature of biology will become more prominent in the near future. For one reason, tools for quantitation have become accessible for many biologists. In addition to -omics technologies, these include the availability of cheap computing power, easily accessible visual simulation software (such as Gepasi [175]), formal languages that can describe biological data and interactions (such as the Systems Biology Markup Language [113]), and integrated approaches to share and access models of many sources (such as the Systems Biology Workbench [222]). For another reason, the consensus in biology has shifted from a reductionist to a more holistic approach to analyze biological systems.

Until recently, extensive knowledge and quantitative data were available for only a few model organisms (notably *Escherichia coli* and *S. cerevisiae*). The current advances in 'Systems Biology' technologies and theoretical frameworks now enable researchers to investigate a much wider range of organisms in greater detail than many had envisioned only 15 years ago. These developments do not only promise a better insight into the peculiarities of one organism under study. Also, they might bring about a clearer understanding of the underlying principles that make life, 'life'.

References

1. **Aderem A.** 2005. Systems biology: its practice and challenges. Cell **121**:511-513.
2. **Affymetrix.** 2001. Statistical algorithms reference guide, Technical report. Affymetrix, Santa Clara, CA USA.
3. **Affymetrix.** 2004. GeneChip expression analysis technical manual. Affymetrix, Santa Clara, CA USA.
4. **Akache B., K. Wu, B. Turcotte.** 2001. Phenotypic analysis of genes encoding yeast zinc cluster proteins. Nucleic Acids Res. **29**:2181-2190.
5. **Albang R., A. Fritz, A. Heinrich, H. Ilgenfritz, D. Maier, C. Wagner, U. Folkers, B. Gerhard, and others.** 2006. Novel lipases and uses thereof. Patent WO/2004/018660.
6. **Albermann K., W. Kemmner, D. Maier, F. Sprcafico, A. Stock, C. Wagner, L. De Boer, R. Meima.** 2007. Novel phospholipases and uses thereof. Patent WO/2003/097825.
7. **Albert R., H. Jeong, A. Barabasi.** 2000. Error and attack tolerance of complex networks. Nature **406**:378-382.
8. **Alwine J. C., D. J. Kemp, G. R. Stark.** 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Natl. Acad. Sci. U.S.A. **74**:5350-5354.
9. **Andersen C. L., J. L. Jensen, T. F. Orntoft.** 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. **64**:5245-5250.
10. **Andersen M. R., M. L. Nielsen, J. Nielsen.** 2008. Metabolic model integration of the bibliome, genome, metabolome and reactome of *Aspergillus niger*. Mol. Syst. Biol. **4**:178.
11. **Andersen M. R., W. Vongsangnak, G. Panagiotou, M. P. Salazar, L. Lehmann, J. Nielsen.** 2008. A trispecies *Aspergillus* microarray: Comparative transcriptomics of three *Aspergillus* species. Proc Natl Acad Sci U S A **105**:4387-4392.
12. **Anderson N. L., N. G. Anderson.** 1998. Proteome and proteomics: new technologies, new concepts, and new words. Electrophoresis **19**:1853-1861.
13. **Ashby W. R.** 1957. An introduction to cybernetics. Chapman & hall Ltd, London.
14. **Auffray C., S. Imbeaud, M. Roux-Rouquié, L. Hood**. 2003. From functional genomics to systems biology: concepts and practices. C. R. Biol. **326**:879-892.
15. **Awasthi V., S. Pandit, P. C. Misra.** 2005. Triton X-100 inhibition of yeast plasma membrane associated NADH-dependent redox activities. J Enzyme Inhib Med Chem **20**:205-209.
16. **Baitaluk M., M. Sedova, A. Ray, A. Gupta.** 2006. BiologicalNetworks: visualization and analysis tool for systems biology. Nucleic Acids Res. **34**:W466-71.
17. **Barabási A., Z. N. Oltvai.** 2004. Network biology: understanding the cell's functional organization. Nat. Rev. Genet. **5**:101-113.
18. **Barbu V., F. Dautry.** 1989. Northern blot normalization with a 28S rRNA oligonucleotide probe. Nucleic Acids Res. **17**:7115.
19. **Baumgartner U., B. Hamilton, M. Piskacek, H. Ruis, H. Rottensteiner.** 1999. Functional analysis of the Zn_2Cys_6 transcription factors Oaf1p and Pip2p. Different roles in fatty acid induction of β-oxidation in *Saccharomyces cerevisiae*. J. Biol. Chem. **274**:22208-22216.
20. **Bedau M. A.** 2003. Artificial life: organization, adaptation and complexity from the bottom up. Trends Cogn. Sci. (Regul. Ed.) **7**:505-512.
21. **Benjamini Y., Y. Hochberg.** 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, series B **57**:289-300.
22. **Bennett J.** 1997. White Paper: Genomics for Filamentous Fungi. Fungal Genet. Biol. **21**:3-7.
23. **Bennett J., L. Lasure.** 1991. Growth Media. In Bennett J, Lasure L (Eds.), More gene manipulations in fungi. Academic Press, San Diego CA USA.
24. **Benson D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, D. L. Wheeler.** 2007. GenBank. Nucleic Acids Res. **35**:D21-5.
25. **Bergmann S., J. Ihmels, N. Barkai.** 2004. Similarities and differences in genome-wide expression data of six organisms. PLoS Biol. **2**:E9.

26. **Bhadauria V., L. Popescu, W. Zhao, Y. Peng.** 2007. Fungal transcriptomics. Microbiol. Res. **162**:285-298.
27. **Bhaumik P., M. K. Koski, T. Glumoff, J. K. Hiltunen, R. K. Wierenga.** 2005. Structural biology of the thioester-dependent degradation and synthesis of fatty acids. Curr. Opin. Struct. Biol. **15**:621-628.
28. **Bohle K., A. Jungebloud, Y. Gocke, A. Dalpiaz, C. Cordes, H. Horn, D. C. Hempel.** 2007. Selection of reference genes for normalisation of specific gene quantification data of *Aspergillus niger*. J. Biotechnol. **132**:353-358.
29. **Bok J. W., D. Hoffmeister, L. A. Maggio-Hall, R. Murillo, J. D. Glasner, N. P. Keller.** 2006. Genomic mining for *Aspergillus* natural products. Chem. Biol. **13**:31-37.
30. **Bork P.**.2000. Powers and pitfalls in sequence analysis: the 70% hurdle. Genome Res. **10**:398-400.
31. **Bothwell J. H. F.** 2006. The long past of systems biology. New Phytol. **170**:6-10.
32. **Bouws H., A. Wattenberg, H. Zorn.** 2008. Fungal secretomes - nature's toolbox for white biotechnology. Appl. Microbiol. Biotechnol. **80**:381-388.
33. **Bower N. I., R. J. Moser, J. R. Hill, S. A. Lehnert.** 2007. Universal reference method for real-time PCR gene expression analysis of preimplantation embryos. BioTechniques **42**:199-206.
34. **Brazma A., H. Parkinson, U. Sarkans, M. Shojatalab, J. Vilo, N. Abeygunawardena, E. Holloway, M. Kapushesky, and others.** 2003. ArrayExpress - a public repository for microarray gene expression data at the EBI. Nucleic Acids Res. **31**:68-71.
35. **Brazma A., P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, and others.** 2001. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat. Genet. **29**:365-371.
36. **Breakspear A., M. Momany.** 2007. The first fifty microarray studies in filamentous fungi. Microbiology **153**:7-15.
37. **Brent M. R.** 2008. Steady progress and recent breakthroughs in the accuracy of automated genome annotation. Nat. Rev. Genet. **9**:62-73.
38. **Brent R.** 2004. A partnership between biology and engineering. Nat. Biotechnol. **22**:1211-1214.
39. **Browse J., C. Somerville.** 1994. Glycerolipids. In Meyerowitz E (Ed.), *Arabidopsis*. Cold Spring Harbour Laboratory Press.
40. **Bruggeman F. J., H. V. Westerhoff.** 2007. The nature of systems biology. Trends Microbiol. **15**:45-50.
41. **Bustin S.** 2002. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. J. Mol. Endocrinol. **29**:23-39.
42. **Carlile M.** 1995. The success of the hypha and mycelium. In Gow N, Gadd G (Eds.), The growing fungus. Chapman & Hall, London.
43. **Carlson M. R. J., B. Zhang, Z. Fang, P. S. Mischel, S. Horvath, S. F. Nelson.** 2006. Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. BMC Genomics **7**:40.
44. **Castrillo J. I., L. A. Zeef, D. C. Hoyle, N. Zhang, A. Hayes, D. C. J. Gardner, M. J. Cornell, J. Petty, and others.** 2007. Growth control of the eukaryote cell: a systems biology study in yeast. J. Biol. **6**:4.
45. **Chang A., M. Scheer, A. Grote, I. Schomburg, D. Schomburg.** 2009. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. Nucleic Acids Res. **37**:D588-92.
46. **Chen K. C., A. Csikasz-Nagy, B. Gyorffy, J. Val, B. Novak, J. J. Tyson.** 2000. Kinetic analysis of a molecular model of the budding yeast cell cycle. Mol. Biol. Cell **11**:369-391.
47. **Chudin E., R. Walker, A. Kosaka, S. X. Wu, D. Rabert, T. K. Chang, D. E. Kreder.** 2002. Assessment of the relationship between signal intensities and transcript concentration for Affymetrix GeneChip arrays. Genome Biol. **3**:RESEARCH0005.
48. **Chung O., C. Tsen.** 1975. Changes in lipid binding and distribution during dough mixing. Cereal Chemistry **52**:533-549.
49. **Cleland C. E., C. F. Chyba.** 2002. Defining 'life'. Orig Life Evol Biosph **32**:387-393.

50. **Conrad P. G., K. H. Nealson.** 2001. A non-earthcentric approach to life detection. Astrobiology **1**:15-24.
51. **Cornish-Bowden A., M. L. Cárdenas.** 2005. Systems biology may work when we learn to understand the parts in terms of the whole. Biochem. Soc. Trans. **33**:516-519.
52. **Cross F. R., V. Archambault, M. Miller, M. Klovstad.** 2002. Testing a mathematical model of the yeast cell cycle. Mol. Biol. Cell **13**:52-70.
53. **Currie C. R.** 2001. A community of ants, fungi, and bacteria: a multilateral approach to studying symbiosis. Annu. Rev. Microbiol. **55**:357-380.
54. **de Graaff L. H., H. C. van den Broeck, A. J. van Ooijen, J. Visser.** 1994. Regulation of the xylanase-encoding *xlnA* gene of *Aspergillus tubigensis*. Mol. Microbiol. **12**:479-490.
55. **de Graaff L., H. van den Broeck, J. Visser.** 1992. Isolation and characterization of the *Aspergillus niger* pyruvate kinase gene. Curr. Genet. **22**:21-27.
56. **De Stefanis V. A., J. G. J. Ponte.** 1969. Simple method for isolation of crude mono- and digalactosyl diglycerides from wheat flour. Biochim. Biophys. Acta **176**:198-201.
57. **De Stefanis V., J. J. Ponte.** 1969. Simple method for isolation of crude mono- and digalactosyl diglycerides from wheat flour. Biochim. Biophys. Acta **176**:198-201.
58. **de Vries R. P., J. Visser.** 1999. Regulation of the feruloyl esterase (*faeA*) gene from *Aspergillus niger*. Appl. Environ. Microbiol. **65**:5500-5503.
59. **de Vries R. P., J. Visser, L. H. de Graaff.** 1999. CreA modulates the XlnR-induced expression on xylose of *Aspergillus niger* genes involved in xylan degradation. Res. Microbiol. **150**:281-285.
60. **DeZwaan T. M., A. M. Carroll, B. Valent, J. A. Sweigard.** 1999. *Magnaporthe grisea* pth11p is a novel plasma membrane protein that mediates appressorium differentiation in response to inductive substrate cues. Plant Cell **11**:2013-2030.
61. **Dheda K., J. F. Huggett, J. S. Chang, L. U. Kim, S. A. Bustin, M. A. Johnson, G. A. W. Rook, A. Zumla.** 2005. The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. Anal. Biochem. **344**:141-143.
62. **Dheda K., J. F. Huggett, S. A. Bustin, M. A. Johnson, G. Rook, A. Zumla.** 2004. Validation of housekeeping genes for normalizing RNA expression in real-time PCR. BioTechniques **37**:112-4, 116, 118-9.
63. **Dirusso C. C., P. N. Black.** 2004. Bacterial long chain fatty acid transport: gateway to a fatty acid-responsive signaling system. J. Biol. Chem. **279**:49563-49566.
64. **Distel B., R. Erdmann, S. J. Gould, G. Blobel, D. I. Crane, J. M. Cregg, G. Dodt, Y. Fujiki, and others.** 1996. A unified nomenclature for peroxisome biogenesis factors. J. Cell Biol. **135**:1-3.
65. **Dormann P., C. Benning.** 2002. Galactolipids rule in seed plants. Trends Plant Sci. **7**:112-118.
66. **Douce R., J. Joyard.** 1980. Plant galactolipids. In Strumf P (Ed.), The Biochemistry of plants. Academic Press, Inc, New York.
67. **Dwight S. S., M. A. Harris, K. Dolinski, C. A. Ball, G. Binkley, K. R. Christie, D. G. Fisk, L. Issel-Tarver, and others.** 2002. *Saccharomyces* Genome Database (SGD) provides secondary gene annotation using the Gene Ontology (GO). Nucleic Acids Res. **30**:69-72.
68. **Eddy S. R.** 2001. Non-coding RNA genes and the modern RNA world. Nat. Rev. Genet. **2**:919-929.
69. **Edgar R., M. Domrachev, A. E. Lash.** 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. **30**:207-210.
70. **Ehinger A., S. H. Denison, G. S. May.** 1990. Sequence, organization and expression of the core histone genes of *Aspergillus nidulans*. Mol. Gen. Genet. **222**:416-424.
71. **Eisen M. B., P. T. Spellman, P. O. Brown, D. Botstein.** 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U.S.A. **95**:14863-14868.
72. **El-Hani C. N.** 2008. Theory-based approaches to the concept of life. J. Biol. Educ. **42**:147-149.
73. **Erdmann R., M. Veenhuis, D. Mertens, W. H. Kunau.** 1989. Isolation of peroxisome-deficient mutants of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. U.S.A. **86**:5419-5423.
74. **Etienne W., M. H. Meyer, J. Peppers, Meyer, R A, Jr.** 2004. Comparison of mRNA gene expression by RT-PCR and DNA microarray. BioTechniques **36**:618-20, 622, 624-6.
75. **Faergeman N. J., C. C. DiRusso, A. Elberger, J. Knudsen, P. N. Black.** 1997. Disruption of the

Saccharomyces cerevisiae homologue to the murine fatty acid transport protein impairs uptake and growth on long-chain fatty acids. J. Biol. Chem. **272:**8531-8538.

76. **Faergeman N. J., P. N. Black, X. D. Zhao, J. Knudsen, C. C. DiRusso.** 2001. The Acyl-CoA synthetases encoded within FAA1 and FAA4 in *Saccharomyces cerevisiae* function as components of the fatty acid transport system linking import, activation, and intracellular utilization. J. Biol. Chem. **276:**37051-37059.
77. **Featherstone D. E., K. Broadie.** 2002. Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network. Bioessays **24**:267-274.
78. **Frey P. A.** 1996. The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. FASEB J. **10:**461-470.
79. **Friedberg I.** 2006. Automated protein function prediction - the genomic challenge. Brief. Bioinformatics **7:**225-242.
80. **Fu L., P. Bounelis, N. Dey, B. L. Browne, R. B. Marchase, D. M. Bedwell.** 1995. The posttranslational modification of phosphoglucomutase is regulated by galactose induction and glucose repression in *Saccharomyces cerevisiae*. J. Bacteriol. **177:**3087-3094.
81. **Futcher B.** 1996. Cyclins and the wiring of the yeast cell cycle. Yeast **12:**1635-1646.
82. **Galperin M. Y., G. R. Cochrane.** 2009. Nucleic Acids Research annual Database Issue and the NAR online Molecular Biology Database Collection in 2009. Nucleic Acids Res. **37**:D1-4.
83. **Gardner M.** 1970. The fantastic combinations of John Conway's new solitaire game "life". Scientific American **223:**120-123.
84. **Gasch A. P., A. M. Moses, D. Y. Chiang, H. B. Fraser, M. Berardini, M. B. Eisen.** 2004. Conservation and evolution of *cis*-regulatory systems in ascomycete fungi. PLoS Biol. **2:**e398.
85. **Ghisla S., C. Thorpe.** 2004. Acyl-CoA dehydrogenases. A mechanistic overview. Eur. J. Biochem. **271**:494-508.
86. **Gibson U. E., C. A. Heid, P. M. Williams.** 1996. A novel method for real time quantitative RT-PCR. Genome Res. **6**:995-1001.
87. **Gilsbach R., M. Kouta, H. Bönisch, M. Brüss.** 2006. Comparison of in vitro and in vivo reference genes for internal standardization of real-time PCR data. BioTechniques **40:**173.
88. **Goodrum L. J., A. Patel, J. F. Leykam, M. J. Kieliszewski.** 2000. Gum arabic glycoprotein contains glycomodules of both extensin and arabinogalactan-glycoproteins. Phytochemistry **54**:99-106.
89. **Goosen T., G. Bloemheuvel, C. Gysler, D. A. de Bie, H. W. van den Broek, K. Swart.** 1987. Transformation of *Aspergillus niger* using the homologous orotidine-5'-phosphate-decarboxylase gene. Curr. Genet. **11**:499-503.
90. **Griffeath D.** 1992. Comment: Randomness in Complex systems. Statistical Science **7**:104-108.
91. **Guillemette T., N. N. M. E. van Peij, T. Goosen, K. Lanthaler, G. D. Robson, C. A. M. J. J. van den Hondel, H. Stam, D. B. Archer.** 2007. Genomic analysis of the secretion stress response in the enzyme-producing cell factory *Aspergillus niger*. BMC Genomics **8:**158.
92. **Güldener U., K. Seong, J. Boddu, S. Cho, F. Trail, J. Xu, G. Adam, H. Mewes, and others.** 2006. Development of a *Fusarium graminearum* Affymetrix GeneChip for profiling fungal gene expression *in vitro* and *in planta*. Fungal Genet. Biol. **43:**316-325.
93. **Gurvitz A., H. Rottensteiner, S. H. Kilpeläinen, A. Hartig, J. K. Hiltunen, M. Binder, I. W. Dawes, B. Hamilton.** 1997. The *Saccharomyces cerevisiae* peroxisomal 2,4-dienoyl-CoA reductase is encoded by the oleate-inducible gene SPS19. J. Biol. Chem. **272:**22140-22147.
94. **Hajri T., N. Abumrad.** 2002. Fatty acid transport across membranes: relevance to nutrition and metabolic pathology. Annu. Rev. Nutr. **22:**383-415.
95. **Harbison C. T., D. B. Gordon, T. I. Lee, N. J. Rinaldi, K. D. Macisaac, T. W. Danford, N. M. Hannett, J. Tagne, and others.** 2004. Transcriptional regulatory code of a eukaryotic genome. Nature **431**:99-104.
96. **Harmon C. M., N. A. Abumrad.** 1993. Binding of sulfosuccinimidyl fatty acids to adipocyte membrane proteins: isolation and amino-terminal sequence of an 88-kD protein implicated in transport of long-chain fatty acids. J. Membr. Biol. **133:**43-49.

97. **Hashimoto H., Y. Kikuchi, Y. Nogi, T. Fukasawa.** 1983. Regulation of expression of the galactose gene cluster in *Saccharomyces cerevisiae*. Isolation and characterization of the regulatory gene GAL4. Mol. Gen. Genet. **191**:31-38.
98. **Hasper A., L. Trindade, D. van der Veen, A. van Ooyen, L. de Graaff.** 2004. Functional analysis of the transcriptional activator XlnR from *Aspergillus niger*. Microbiology **150**:1367-1375.
99. **Hasper L.** Function and mode of regulation of the transcriptional activator XlnR from *Aspergillus*. 2003. Ph.D. thesis. Wageningen University.
100. **Hatzikirou H., A. Deutsch.** 2008. Cellular automata as microscopic models of cell migration in heterogeneous environments. Curr. Top. Dev. Biol. **81**:401-434.
101. **Hawksworth D. L.** 2001. The magnitude of fungal diversity: the 1.5 million species estimate revisited. Mycological Research **105**:1422-1432.
102. **Hayes J. D., R. C. Strange.** 2000. Glutathione S-transferase polymorphisms and their biological consequences. Pharmacology **61**:154-166.
103. **Hettema E. H., C. W. van Roermund, B. Distel, M. van den Berg, C. Vilela, C. Rodrigues-Pousada, R. J. Wanders, H. F. Tabak.** 1996. The ABC transporter proteins Pat1 and Pat2 are required for import of long-chain fatty acids into peroxisomes of *Saccharomyces cerevisiae*. EMBO J. **15**:3813-3822.
104. **Hibbeler S., J. P. Scharsack, S. Becker.** 2008. Housekeeping genes for quantitative expression studies in the three-spined stickleback *Gasterosteus aculeatus*. BMC Mol. Biol. **9**:18.
105. **Hibbett D. S., M. Binder, J. F. Bischoff, M. Blackwell, P. F. Cannon, O. E. Eriksson, S. Huhndorf, T. James, and others.** 2007. A higher-level phylogenetic classification of the Fungi. Mycol. Res. **111**:509-547.
106. **Higuchi R., G. Dollinger, P. S. Walsh, R. Griffith.** 1992. Simultaneous amplification and detection of specific DNA sequences. Biotechnology (N.Y.) **10**:413-417.
107. **Hiltunen J. K., A. M. Mursula, H. Rottensteiner, R. K. Wierenga, A. J. Kastaniotis, A. Gurvitz.** 2003. The biochemistry of peroxisomal β-oxidation in the yeast *Saccharomyces cerevisiae*. FEMS Microbiol. Rev. **27**:35-64.
108. **Hoffmeister D., N. P. Keller.** 2007. Natural products of filamentous fungi: enzymes, genes, and their regulation. Nat Prod Rep **24**:393-416.
109. **Hong E. L., R. Balakrishnan, Q. Dong, K. R. Christie, J. Park, G. Binkley, M. C. Costanzo, S. S. Dwight, and others.** 2008. Gene Ontology annotations at SGD: new data sources and annotation methods. Nucleic Acids Res. **36**:D577-81.
110. **Horak C. E., M. Snyder.** 2002. ChIP-chip: a genomic approach for identifying transcription factor binding sites. Meth. Enzymol. **350**:469-483.
111. **Hsu K., G. Lee, J. Shaw.** 2008. Promoter analysis and differential expression of the *Candida rugosa* lipase gene family in response to culture conditions. J. Agric. Food Chem. **56**:1992-1998.
112. **Hube B., F. Stehr, M. Bossenz, A. Mazur, M. Kretschmar, W. Schäfer.** 2000. Secreted lipases of *Candida albicans*: cloning, characterisation and expression analysis of a new gene family with at least ten members. Arch. Microbiol. **174**:362-374.
113. **Hucka M., A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, and others.** 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics **19**:524-531.
114. **Huggett J., K. Dheda, S. Bustin, A. Zumla.** 2005. Real-time RT-PCR normalisation; strategies and considerations. Genes Immun. **6**:279-284.
115. **Hughes T. R., M. D. Robinson, N. Mitsakakis, M. Johnston.** 2004. The promise of functional genomics: completing the encyclopedia of a cell. Curr. Opin. Microbiol. **7**:546-554.
116. **Hughes T. R., M. J. Marton, A. R. Jones, C. J. Roberts, R. Stoughton, C. D. Armour, H. A. Bennett, E. Coffey, and others.** 2000. Functional discovery via a compendium of expression profiles. Cell **102**:109-126.
117. **Hulsen T., M. A. Huynen, J. de Vlieg, P. M. A. Groenen.** 2006. Benchmarking ortholog identification methods using functional genomics data. Genome Biol. **7**:R31.
118. **Hynes M. J.** 1977. Induction of the acetamidase of *Aspergillus nidulans* by acetate metabolism. J.

Bacteriol. **131**:770-775.

119. **Hynes M. J., O. W. Draht, M. A. Davis.** 2002. Regulation of the *acuF* gene, encoding phosphoenolpyruvate carboxykinase in the filamentous fungus *Aspergillus nidulans*. J. Bacteriol. 184:183-190.
120. **Hynes M. J., S. L. Murray, A. Duncan, G. S. Khew, M. A. Davis.** 2006. Regulatory genes controlling fatty acid catabolism and peroxisomal functions in the filamentous fungus *Aspergillus nidulans*. Eukaryotic Cell **5**:794.
121. **Hynes M. J., S. L. Murray, G. S. Khew, M. A. Davis.** 2008. Genetic analysis of the role of peroxisomes in the utilization of acetate and fatty acids in *Aspergillus nidulans*. Genetics **178**:1355-1369.
122. **Ideker T., V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, and others.** 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science **292**:929-934.
123. **Imbeaud S., C. Auffray.** 2005. 'The 39 steps' in gene expression profiling: critical issues and proposed best practices for microarray experiments. Drug Discov. Today **10**:1175-1182.
124. **Irizarry R., B. Bolstad, F. Collin, L. Cope, B. Hobbs, T. Speed.** 2003. Summaries of Affymetrix Genechip probe level data. Nucleic Acids Res. **31**:e15.
125. **Jafari P., F. Azuaje.** 2006. An assessment of recently published gene expression data analyses: reporting experimental design and statistical factors. BMC Med Inform Decis Mak **6**:27.
126. **Johnson D. R., L. J. Knoll, D. E. Levin, J. I. Gordon.** 1994. *Saccharomyces cerevisiae* contains four fatty acid activation (FAA) genes: an assessment of their role in regulating protein N-myristoylation and cellular lipid metabolism. J. Cell Biol. **127**:751-762.
127. **Johnston M.** 1987. A model fungal gene regulatory mechanism: the GAL genes of *Saccharomyces cerevisiae*. Microbiol. Rev. **51**:458-476.
128. Joint Genome Institute website, http://genome.jgi-psf.org/Aspni5/Aspni5.home.html.
129. **Jones G. W., P. Hooley, S. M. Farrington, S. G. Shawcross, L. A. Iwanejko, P. Strike.** 1999. Cloning and characterisation of the *sagA* gene of *Aspergillus nidulans*: a gene which affects sensitivity to DNA-damaging agents. Mol. Gen. Genet. **261**:251-258.
130. **Jørgensen T. R., T. Goosen, C. A. M. J. J. V. D. Hondel, A. F. J. Ram, J. J. L. Iversen.** 2009. Transcriptomic comparison of *Aspergillus niger* growing on two different sugars reveals coordinated regulation of the secretory pathway. BMC Genomics **10**:44.
131. **Kamini N., J. Geraldine Sandana Mala, R. Purvanakrishnan.** 1997. Production and characterization of an extracellular lipase from *Aspergillus niger*. Indian Journal of Microbiology **37**:85-89.
132. **Karpichev I. V., G. M. Small.** 1998. Global regulatory functions of Oaf1p and Pip2p (Oaf2p), transcription factors that regulate genes encoding peroxisomal proteins in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **18**:6560-6570.
133. **Kasuga T., N. L. Glass.** 2008. Dissecting colony development of *Neurospora crassa* using mRNA profiling and comparative genomics approaches. Eukaryotic Cell **7**:1549-1564.
134. **Katz M. E., M. J. Hynes.** 1989. Isolation and analysis of the acetate regulatory gene, *facB*, from *Aspergillus nidulans*. Mol. Cell. Biol. **9**:5696-5701.
135. **Kell D. B., S. G. Oliver.** 2004. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. Bioessays **26**:99-105.
136. **Kiel J. A. K. W., M. Veenhuis, I. J. van der Klei.** 2006. PEX genes in fungal genomes: common, rare or redundant. Traffic **7**:1291-1303.
137. **Kim Y., M. P. Nandakumar, M. R. Marten.** 2007. Proteomics of filamentous fungi. Trends Biotechnol. **25**:395-400.
138. **Kim Y., M. P. Nandakumar, M. R. Marten.** 2008. The state of proteome profiling in the fungal genus *Aspergillus*. Brief Funct Genomic Proteomic **7**:87-94.
139. **Kinoshita K., M. Takano, T. Koseki, K. Ito, K. Iwano.** 1995. Cloning of the *xynB* gene encoding xylanase B from *Aspergillus niger* and its expression in *Aspergillus kawachii*. J. Ferment. Bioeng.

79:422-428.

140. **Kitano H..** 2002. Systems biology: a brief overview. Science **295**:1662-1664.
141. **Klein H. P.** 1992. The Viking biology experiments: epilogue and prologue. Orig Life Evol Biosph **21**:255-261.
142. **Kleppe K., E. Ohtsuka, R. Kleppe, I. Molineux, H. G. Khorana.** 1971. Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases. J. Mol. Biol. **56**:341-361.
143. **Klevecz R. R., C. M. Li, I. Marcus, P. H. Frankel.** 2008. Collective behavior in gene regulation: the cell is an oscillator, the cell cycle a developmental process. FEBS J. **275**:2372-2384.
144. **Knijnenburg T. A., L. F. A. Wessels, M. J. T. Reinders.** 2008. Combinatorial influence of environmental parameters on transcription factor activity. Bioinformatics **24**:i172-81.
145. **Korzeniewski B.** 2001. Cybernetic formulation of the definition of life. J. Theor. Biol. **209**:275-286.
146. **Korzeniewski B.** 2004. Confrontation of the cybernetic definition of a living individual with the real world. Acta Biotheoretica **53**:1-28.
147. **Kulkarni R. D., M. R. Thon, H. Pan, R. A. Dean.** 2005. Novel G-protein-coupled receptor-like proteins in the plant pathogenic fungus *Magnaporthe grisea*. Genome Biol. **6**:R24.
148. **Kunau W. H., V. Dommes, H. Schulz.** 1995. β-Oxidation of fatty acids in mitochondria, peroxisomes, and bacteria: a century of continued progress. Prog. Lipid Res. **34**:267-342.
149. **Kusters-van Someren M. A., J. A. Harmsen, H. C. Kester, J. Visser.** 1991. Structure of the *Aspergillus niger pelA* gene and its expression in *Aspergillus niger* and *Aspergillus nidulans*. Curr. Genet. **20**:293-299.
150. **Lashkari D. A., J. L. DeRisi, J. H. McCusker, A. F. Namath, C. Gentile, S. Y. Hwang, P. O. Brown, R. W. Davis.** 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc. Natl. Acad. Sci. U.S.A. **94**:13057-13062.
151. **Lazarow P. B.** 2003. Peroxisome biogenesis: advances and conundrums. Curr. Opin. Cell Biol. **15**:489-497.
152. **Lee S., M. Jo, J. Lee, S. S. Koh, S. Kim.** 2007. Identification of novel universal housekeeping genes by statistical analysis of microarray data. J. Biochem. Mol. Biol. **40**:226-231.
153. **Levin A. M., R. P. de Vries, A. Conesa, C. de Bekker, M. Talon, H. H. Menke, N. N. M. E. van Peij, H. A. B. Wösten.** 2007. Spatial differentiation in the vegetative mycelium of *Aspergillus niger*. Eukaryotic Cell **6**:2311-2322.
154. **Lin S. J., V. C. Culotta.** 1996. Suppression of oxidative damage by *Saccharomyces cerevisiae* ATX2, which encodes a manganese-trafficking protein that localizes to Golgi-like vesicles. Mol. Cell. Biol. **16**:6303-6312.
155. **Lisacek F., R. D. Appel.** 2007. Systems Biology: a loose definition. Proteomics **7**:825-827.
156. **Lockhart D. J., H. Dong, M. C. Byrne, M. T. Follettie, M. V. Gallo, M. S. Chee, M. Mittmann, C. Wang, and others.** 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. **14**:1675-1680.
157. **Lohr D., P. Venkov, J. Zlatanova.** 1995. Transcriptional regulation in the yeast GAL gene family: a complex genetic network. FASEB J. **9**:777-787.
158. **Lucas J. D., A. Dominquez, S. Valenciano, G. Turner, F. Laborda.** 1999. The *acuH* gene of *Aspergillus nidulans*, required for growth on acetate and long-chain fatty acids, encodes a putative homologue of the mammalian carnitine/acylcarnitine carrier. Arch. Microbiol. **171**:386-396.
159. **Lucas J. D., S. Valenciano, A. Dominquez, G. Turner, F. Laborda.** 1997. Characterization of oleate-nonutilizing mutants of *Aspergillus nidulans* isolated by the 3-amino-1,2,4-triazole positive selection method. Arch. Microbiol. **168**.
160. **Luisi P. L.** 1998. About various definitions of life. Orig Life Evol Biosph **28**:613-622.
161. **Lutfalla G., G. Uze.** 2006. Performing quantitative reverse-transcribed polymerase chain reaction experiments. Meth. Enzymol. **410**:386-400.
162. **Lyng M. B., A. Laenkholm, N. Pallisgaard, H. J. Ditzel.** 2008. Identification of genes for

normalization of real-time RT-PCR data in breast carcinomas. BMC Cancer **8**:20.

163. **MacKenzie D. A., T. Guillemette, H. Al-Sheikh, A. J. Watson, D. J. Jeenes, P. Wongwathanarat, N. S. Dunn-Coleman, N. van Peij, and others.** 2005. UPR-independent dithiothreitol stress-induced genes in *Aspergillus niger*. Mol. Genet. Genomics **274**:410-418.
164. **MacPherson S., M. Larochelle, B. Turcotte.** 2006. A fungal family of transcriptional regulators: the zinc cluster proteins. Microbiol. Mol. Biol. Rev. **70**:583-604.
165. **Maggio-Hall L. A., P. Lyne, J. A. Wolff, N. P. Keller.** 2008. A single acyl-CoA dehydrogenase is required for catabolism of isoleucine, valine and short-chain fatty acids in *Aspergillus nidulans*. Fungal Genet. Biol. **45**:180-189.
166. **Maggio-Hall L., N. Keller.** 2004. Mitochondrial β-oxidation in *Aspergillus nidulans*. Molecular Microbiology **54**:1173-1185.
167. **Magnuson J. K., L. L. Lasure.** 2004. Organic acid production by Filamentous Fungi. In Lange J, Lange L (Eds.), Advances in Fungal Biotechnology for Industry, Agriculture, and Medicine, Kluwer Academic/Plenum Publishers.
168. **Maicas S., I. Moreno, A. Nieto, M. Gómez, R. Sentandreu, E. Valentín.** 2005. In Silico Analysis for Transcription Factors With $Zn(II)_2Cys_6$ Binuclear Cluster DNA-Binding Domains in *Candida albicans*. Comp. Funct. Genomics **6**:345-356.
169. **Margaritis T., F. C. P. Holstege.** 2008. Poised RNA polymerase II gives pause for thought. Cell **133**:581-584.
170. **Marmorstein R., M. Carey, M. Ptashne, S. C. Harrison.** 1992. DNA recognition by GAL4: structure of a protein-DNA complex. Nature **356**:408-414.
171. **Marmorstein R., S. C. Harrison.** 1994. Crystal structure of a PPR1-DNA complex: DNA recognition by proteins containing a Zn_2Cys_6 binuclear cluster. Genes Dev. **8**:2504-2512.
172. **Martens-Uzunova E.**S. 2008. Assessment of the pectinolytic network of *Aspergillus niger* by functional transcriptomics. Insights from the transcriptome. Ph.D. thesis. Wageningen University.
173. **Martens-Uzunova E. S., J. S. Zandleven, J. A. Benen, H. Awad, H. J. Kools, G. Beldman, A. G. Voragen, J. A. Van den Berg, and others.** 2006. A new group of exo-acting family 28 glycoside hydrolases of *Aspergillus niger* that are involved in pectin degradation. Biochem. J. **400**:43-52.
174. **Mata J., M. Cohn.** 2007. Cellular automata-based modeling program: synthetic immune system. Immunol. Rev. **216**:198-212.
175. **Mendes P.** 1997. Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3. Trends Biochem. Sci. **22**:361-363.
176. **Meyer V., R. A. Damveld, M. Arentshorst, U. Stahl, C. A. van den Hondel, A. F. Ram.** 2007. Survival in the presence of antifungals: genome-wide expression profiling of *Aspergillus niger* in response to sublethal concentrations of caspofungin and fenpropimorph. J. Biol. Chem. **282**:32935-32948.
177. **Misra R. V., R. S. P. Horler, W. Reindl, I. I. Goryanin, G. H. Thomas.** 2005. EchoBASE: an integrated post-genomic database for *Escherichia coli*. Nucleic Acids Res. **33**:D329-33.
178. **Morey J. S., J. C. Ryan, F. M. Van Dolah.** 2006. Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR. Biol Proced Online **8**:175-193.
179. **Mueller O., K. Hahnenberger, M. Dittmann, H. Yee, R. Dubrow, R. Nagle, D. Ilsley.** 2000. A microfluidic system for high-speed reproducible DNA sizing and quantitation. Electrophoresis **21**:128-134.
180. **Mullaney E. J., J. E. Hamer, K. A. Roberti, M. M. Yelton, W. E. Timberlake.** 1985. Primary structure of the *trpC* gene from *Aspergillus nidulans.* Mol. Gen. Genet. **199**:37-45.
181. **Muller P., H. Janovjak, A. Miserez, Z. Dobbie.** 2002. Processing of gene expression data generated by quantitative real-time RT-PCR. BioTechniques **32**:1372-1379.
182. **Mullis K. B.** 1987. Process for amplifying, detecting, and/or cloning nucleic acid sequences using a thermostable enzyme. Patent US 4,683,202.
183. **Murone M., V. Simanis.** 1996. The fission yeast *dma1* gene is a component of the spindle

assembly checkpoint, required to prevent septum formation and premature exit from mitosis if spindle function is compromised. Embo J **15**:6605-6616.

184. **Murphy D. J.** 2001. The biogenesis and functions of lipid bodies in animals, plants and microorganisms. Prog. Lipid Res. **40**:325-438.
185. **Nakagawa T., T. Imanaka, M. Morita, K. Ishiguro, H. Yurimoto, A. Yamashita, N. Kato, Y. Sakai.** 2000. Peroxisomal membrane protein Pmp47 is essential in the metabolism of middle-chain fatty acid in yeast peroxisomes and Is associated with peroxisome proliferation. J. Biol. Chem. **275**:3455-3461.
186. **National Center for Biotechnology Information** website, http://www.ncbi.nlm.nih.gov/
187. **Nielsen H., J. Engelbrecht, S. Brunak, G. von Heijne.** 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10**:1-6.
188. **Nierman W. C., A. Pain, M. J. Anderson, J. R. Wortman, H. S. Kim, J. Arroyo, M. Berriman, K. Abe, and others.** 2005. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. Nature **438**:1151-1156.
189. **Nikolaev I., F. Lenouvel, B. Felenbok.** 1999. Unique DNA binding specificity of the binuclear zinc AlcR activator of the ethanol utilization pathway in *Aspergillus nidulans*. J. Biol. Chem. **274**:9795-9802.
190. **Noël J., B. Turcotte.** 1998. Zinc cluster proteins Leu3p and Uga3p recognize highly related but distinct DNA targets. J. Biol. Chem. **273**:17463-17468.
191. **Nygard A., C. B. Jørgensen, S. Cirera, M. Fredholm.** 2007. Selection of reference genes for gene expression studies in pig tissues using SYBR green qPCR. BMC Mol. Biol. **8**:67.
192. **O'Malley M. A., J. Dupré.** 2005. Fundamental issues in systems biology. Bioessays **27**:1270-1276.
193. **Oliveira J. M., D. van der Veen, L. H. de Graaff, L. Qin.** 2008. Efficient cloning system for construction of gene silencing vectors in *Aspergillus niger*. Appl. Microbiol. Biotechnol. **80**:917-924.
194. **Oparin A.** 1924. The Origin of Life. Izd. Moskovhii RabochiI, Moskow
195. **Patterson H. D., R. Thompson.** 1971. Recovery of inter-block information when block sizes are unequal. Biometrika **58**:545-554.
196. **Pel H. J., J. H. de Winde, D. B. Archer, P. S. Dyer, G. Hofmann, P. J. Schaap, G. Turner, R. P. de Vries, and others.** 2007. Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. Nat. Biotechnol. **25**:221-231.
197. **Peng G., J. E. Hopper.** 2000. Evidence for Gal3p's cytoplasmic location and Gal80p's dual cytoplasmic-nuclear location implicates new mechanisms for controlling Gal4p activity in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **20**:5140-5148.
198. **Pfaffl M.** 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. **2001**:9.
199. **Pfaffl M. W., A. Tichopad, C. Prgomet, T. P. Neuvians.** 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper -Excel-based tool using pair-wise correlations. Biotechnol. Lett. **26**:509-515.
200. **Pieterse B., R. H. Jellema, M. J. van der Werf.** 2006. Quenching of microbial samples for increased reliability of microarray data. J. Microbiol. Methods **64**:207-216.
201. **Platt A., R. J. Reece.** 1998. The yeast galactose genetic switch is mediated by the formation of a Gal4p-Gal80p-Gal3p complex. EMBO J. **17**:4086-4091.
202. **Poirier Y., V. D. Antonenkov, T. Glumoff, J. K. Hiltunen.** 2006. Peroxisomal β-oxidation - a metabolic pathway with multiple functions. Biochim. Biophys. Acta **1763**:1413-1426.
203. **Pokorny D., J. Friedrich, A. Cimerman.** 1994. Effect of nutritional factors on lipase biosynthesis by *Aspergillus niger*. Biotechnology letters **16**:363-366.
204. **Pontecorvo G., J. A. Roper, L. M. Hemmons, K. D. Macdonald, A. W. Bufton.** 1953. The genetics of *Aspergillus nidulans*. Adv. Genet. **5**:141-238.
205. **Provenzano M., S. Mocellin.** 2007. Complementary techniques: validation of gene expression data by quantitative real time PCR. Adv. Exp. Med. Biol. **593**:66-73.
206. **Ptashne M., A. Gann.** 2003. Genes and Signals. Cold Spring Harbour Laboratory Press, Cold

Spring Harbour USA.

207. **Punt P. J., N. van Biezen, A. Conesa, A. Albers, J. Mangnus, C. van den Hondel.** 2002. Filamentous fungi as cell factories for heterologous protein production. Trends Biotechnol. **20**:200-206.

208. **Punt P., C. Van den Hondel**. 1992. Analysis of transcription control sequences in filamentous fungi. In: Stahl U, Tudzynscki P (Eds.), Proc. EMBO Workshop on Molecular Biology of Filamentous Fungi. VCH, Weinheim.

209. **Qin L. X., R. P. Beyer, F. N. Hudson, N. J. Linford, D. E. Morris, K. F. Kerr.** 2006. Evaluation of methods for oligonucleotide array data via quantitative real-time PCR. BMC Bioinformatics **7**:23.

210. **Quinn G. P., M. J. Keough.** 2007. Experimental Design and Data Analysis for Biologists . Cambridge University Press.

211. **Reece R. J., M. Ptashne.** 1993. Determinants of binding-site specificity among yeast Cys_6 zinc cluster proteins. Science **261**:909-911.

212. **Ren B., F. Robert, J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, and others.** 2000. Genome-wide location and function of DNA binding proteins. Science **290**:2306-2309.

213. **Robinson T. L., I. A. Sutherland, J. Sutherland.** 2007. Validation of candidate bovine reference genes for use with real-time PCR. Vet. Immunol. Immunopathol. **115**:160-165.

214. **Rohde J. R., J. Trinh, I. Sadowski.** 2000. Multiple signals regulate GAL transcription in yeast. Mol. Cell. Biol. **20**:3880-3886.

215. **Rottensteiner H., A. J. Kal, M. Filipits, M. Binder, B. Hamilton, H. F. Tabak, H. Ruis.** 1996. Pip2p: a transcriptional regulator of peroxisome proliferation in the yeast *Saccharomyces cerevisiae*. EMBO J. **15**:2924-2934.

216. **Rottensteiner H., L. Palmieri, A. Hartig, B. Hamilton, H. Ruis, R. Erdmann, A. Gurvitz.** 2002. The peroxisomal transporter gene ANT1 is regulated by a deviant oleate response element (ORE): characterization of the signal for fatty acid induction. Biochem. J. **365**:109-117.

217. **Rottensteiner H., L. Wabnegger, R. Erdmann, B. Hamilton, H. Ruis, A. Hartig, A. Gurvitz.** 2003. *Saccharomyces cerevisiae* PIP2 mediating oleic acid induction and peroxisome proliferation is regulated by Adr1p and Pip2p-Oaf1p. J. Biol. Chem. **278**:27605-27611.

218. **Rozen S., H. Skaletsky**. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. **132**:365-386.

219. **Ruiz-Mirazo K., J. Peretó, A. Moreno.** 2004. A universal definition of life: autonomy and open-ended evolution. Orig Life Evol Biosph **34**:323-346.

220. **Saiki R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, N. Arnheim.** 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230**:1350-1354.

221. **Salit M., A. Kimmel, O. Brian.** 2006. Standards in gene expression microarray experiments. Meth. Enzymol. Volume **411**:63.

222. **Sauro H. M., M. Hucka, A. Finney, C. Wellock, H. Bolouri, J. Doyle, H. Kitano.** 2003. Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. OMICS **7**:355-372.

223. **Sawers R. J. H., C. Gutjahr, U. Paszkowski.** 2008. Cereal mycorrhiza: an ancient symbiosis in modern agriculture. Trends Plant Sci. **13**:93-97.

224. **Schaffer J. E., H. F. Lodish.** 1994. Expression cloning and characterization of a novel adipocyte long chain fatty acid transport protein. Cell **79**:427-436.

225. **Schena M., D. Shalon, R. W. Davis, P. O. Brown.** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270**:467-470.

226. **Schjerling C. K., R. Hummel, J. K. Hansen, C. Borsting, J. M. Mikkelsen, K. Kristiansen, J. Knudsen.** 1996. Disruption of the gene encoding the acyl-CoA-binding protein (ACB1) perturbs acyl-CoA metabolism in *Saccharomyces cerevisiae*. J. Biol. Chem. **271**:22514-22521.

227. **Schjerling P., S. Holmberg.** 1996. Comparative amino acid sequence analysis of the Cys_6 zinc cluster family of transcriptional regulators. Nucleic Acids Res. **24**:4599-4607.

228. **Schliebs W., C. Würtz, W. Kunau, M. Veenhuis, H. Rottensteiner.** 2006. A eukaryote without catalase-containing microbodies: *Neurospora crassa* exhibits a unique cellular distribution of its four catalases. Eukaryotic Cell **5**:1490-1502.
229. **Schliebs W., W. Kunau.** 2004. Peroxisome membrane biogenesis: the stage is set. Curr. Biol. **14**:R397-9.
230. **Schroeder A., O. Mueller, S. Stocker, R. Salowsky, M. Leiber, M. Gassmann, S. Lightfoot, W. Menzel, and others.** 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol. Biol. **7**:3.
231. **Shalon D., S. J. Smith, P. O. Brown.** 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res. **6**:639-645.
232. **Shannon P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and others.** 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. **13**:2498-2504.
233. **Shearstone J. R., N. E. Allaire, J. Campos-Rivera, S. Rao, S. Perrin.** 2006. Accurate and precise transcriptional profiles from 50 pg of total RNA or 100 flow-sorted primary lymphocytes. Genomics **88**:111-121.
234. **Shipe, WF, Jr.** 1950. A study of the relative specificity of lipases produced by *Penicillium roqueforti* and *Aspergillus niger*. Arch Biochem **30**:165-179.
235. **Shroff R. A., S. M. O'Connor, M. J. Hynes, R. A. Lockington, J. M. Kelly.** 1997. Null alleles of *creA*, the regulator of carbon catabolite repression in *Aspergillus nidulans*. Fungal Genet. Biol. **22**:28-38.
236. **Skoneczny M., A. Chełstowska, J. Rytka.** 1988. Study of the co-induction by fatty acids of catalase A and acyl-CoA oxidase in standard and mutant *Saccharomyces cerevisiae* strains. Eur. J. Biochem. **174**:297-302.
237. **Smith J. J., M. Marelli, R. H. Christmas, F. J. Vizeacoumar, D. J. Dilworth, T. Ideker, T. Galitski, K. Dimitrov, and others.** 2002. Transcriptome profiling to identify genes involved in peroxisome assembly and function. J. Cell Biol. **158**:259-271.
238. **Smith R. D., B. Brown, P. Ikonomi, A. N. Schechter.** 2003. Exogenous reference RNA for normalization of real-time quantitative PCR. BioTechniques **34**:88-91.
239. **Smyth G. K.** 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol **3**:Article3.
240. **Snyder M., M. Gerstein.** 2003. Genomics. Defining genes in the genomics era. Science **300**:258-260.
241. **Stahl A., R. E. Gimeno, L. A. Tartaglia, H. F. Lodish.** 2001. Fatty acid transport proteins: a current view of a growing family. Trends Endocrinol. Metab. **12**:266-273.
242. **Stemple C. J., M. Davis, M. Hynes.** 1998. The *facC* gene of *Aspergillus nidulans* encodes an acetate-inducible carnitine acetyltransferase. J. Bacteriol. **180**:6242-6251.
243. **Stremmel W., G. Strohmeyer, F. Borchard, S. Kochwa, P. D. Berk.** 1985. Isolation and partial characterization of a fatty acid binding protein in rat liver plasma membranes. Proc. Natl. Acad. Sci. U.S.A. **82**:4-8.
244. **Stryer L.** 1995. Biochemistry 4th ed.. W. H. Freeman and Company, New York
245. **Subramani S.** 1998. Components involved in peroxisome import, biogenesis, proliferation, turnover, and movement. Physiol. Rev. **78**:171-188.
246. **Suzuki T., P. J. Higgins, D. R. Crawford.** 2000. Control selection for RNA quantitation. BioTechniques **29**:332-337.
247. **Swaminathan K., P. Flynn, R. J. Reece, R. Marmorstein.** 1997. Crystal structure of a PUT3-DNA complex reveals a novel mechanism for DNA recognition by a protein containing a Zn_2Cys_6 binuclear cluster. Nat. Struct. Biol. **4**:751-759.
248. **Swoboda R. K., G. Bertram, S. Delbrück, J. F. Ernst, N. A. Gow, G. W. Gooday, A. J. Brown.** 1994. Fluctuations in glycolytic mRNA levels during morphogenesis in *Candida albicans* reflect underlying changes in growth and are not a response to cellular dimorphism. Mol. Microbiol. **13**:663-672.

249. **Tan H., K. Okazaki, I. Kubota, T. Kamiryo, H. Utiyama.** 1990. A novel peroxisomal nonspecific lipid-transfer protein from *Candida tropicalis*. Gene structure, purification and possible role in β-oxidation. Eur. J. Biochem. **190**:107-112.
250. **Tanay A., A. Regev, R. Shamir.** 2005. Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. Proc. Natl. Acad. Sci. U.S.A. **102**:7203-7208.
251. **The Gene Ontology Consortium.** 2008. The Gene Ontology project in 2008. Nucleic Acids Res. **36**:D440-4.
252. **Thellin O., W. Zorzi, B. Lakaye, B. De Borman, B. Coumans, G. Hennen, T. Grisar, A. Igout, and others.** 1999. Housekeeping genes as internal standards: use and limits. J. Biotechnol. **75**:291-295.
253. **Tichopad A., M. Dilger, G. Schwarz, M. Pfaffl.** 2003. Standardized determination of real-time PCR efficiency from a single reaction set-up. Nucleic Acids Res. **31**:e122.
254. **Timson D. J., H. C. Ross, R. J. Reece.** 2002. Gal3p and Gal1p interact with the transcriptional repressor Gal80p to form a complex of 1:1 stoichiometry. Biochem. J. **363**:515-520.
255. **Todd R. B., A. Andrianopoulos.** 1997. Evolution of a fungal regulatory gene family: the $Zn(II)_2Cys_6$ binuclear cluster DNA binding motif. Fungal Genet. Biol. **21**:388-405.
256. **Todd R. B., A. Andrianopoulos, M. A. Davis, M. J. Hynes.** 1998. FacB, the *Aspergillus nidulans* activator of acetate utilization genes, binds dissimilar DNA sequences. EMBO J. **17**:2042-2054.
257. **Todd R. B., J. M. Kelly, M. A. Davis, M. J. Hynes.** 1997. Molecular characterization of mutants of the acetate regulatory gene *facB* of *Aspergillus nidulans*. Fungal Genet. Biol. **22**:92-102.
258. **Todd R. B., R. L. Murphy, H. M. Martin, J. A. Sharp, M. A. Davis, M. E. Katz, M. J. Hynes.** 1997. The acetate regulatory gene *facB* of *Aspergillus nidulans* encodes a $Zn(II)_2Cys_6$ transcriptional activator. Mol. Gen. Genet. **254**:495-504.
259. **Tuma R. S., M. P. Beaudet, X. Jin, L. J. Jones, C. Y. Cheung, S. Yue, V. L. Singer.** 1999. Characterization of SYBR Gold nucleic acid gel stain: a dye optimized for use with 300-nm ultraviolet transilluminators. Anal. Biochem. **268**:278-288.
260. **Valenciano S., J. D. Lucas, A. Pedregosa, I. Monistrol, F. Laborda.** 1996. Induction of β-oxidation enzymes and microbody proliferation in *Aspergillus nidulans*. Arch. Microbiol. **166**:336-341.
261. **Vallee B. L., J. E. Coleman, D. S. Auld.** 1991. Zinc fingers, zinc clusters, and zinc twists in DNA-binding protein domains. Proc. Natl. Acad. Sci. U.S.A. **88**:999-1003.
262. **van de Vondervoort P. J. I., S. M. J. Langeveld, J. Visser, N. N. M. E. van Peij, H. J. Pel, C. A. M. J. J. van den Hondel, A. F. J. Ram.** 2007. Identification of a mitotic recombination hotspot on chromosome III of the asexual fungus *Aspergillus niger* and its possible correlation with elevated basal transcription. Curr. Genet. **52**:107-114.
263. **van den Berg R. A., H. C. J. Hoefsloot, J. A. Westerhuis, A. K. Smilde, M. J. van der Werf.** 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics **7**:142.
264. **van der Veen D., J. M. Oliveira, W. A. M. van den Berg, L. H. de Graaff.** 2009. Variance components analysis reveals contribution of sample processing to transcript variation. Appl. Environ. Microbiol. **75**:doi:10.1128/AEM.02270-08
265. **van der Werf M. J., K. M. Overkamp, B. Muilwijk, L. Coulier, T. Hankemeier.** 2007. Microbial metabolomics: toward a platform with full metabolome coverage. Anal. Biochem. **370**:17-25.
266. **van der Zand A., I. Braakman, H. J. Geuze, H. F. Tabak.** 2006. The return of the peroxisome. J. Cell. Sci. **119**:989-994.
267. **van Hartingsveldt W., I. E. Mattern, C. M. van Zeijl, P. H. Pouwels, C. A. van den Hondel.** 1987. Development of a homologous transformation system for *Aspergillus niger* based on the *pyrG* gene. Mol. Gen. Genet. **206**:71-75.
268. **van Noort V., B. Snel, M. A. Huynen.** 2004. The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple model. EMBO Rep. **5**:280-284.
269. **van Peij N. N., J. Brinkmann, M. Vrsanská, J. Visser, L. H. de Graaff.** 1997. β-Xylosidase activity, encoded by *xlnD*, is essential for complete hydrolysis of xylan by *Aspergillus niger* but not

for induction of the xylanolytic enzyme spectrum. Eur. J. Biochem. **245**:164-173.

270. **van Peij N. N., J. Visser, L. H. de Graaff.** 1998. Isolation and analysis of *xlnR*, encoding a transcriptional activator co-ordinating xylanolytic expression in *Aspergillus niger*. Mol. Microbiol. 27:131-142.
271. **van Peij N. N., M. M. Gielkens, R. P. de Vries, J. Visser, L. H. de Graaff.** 1998. The transcriptional activator XlnR regulates both xylanolytic and endoglucanase gene expression in *Aspergillus niger*. Appl. Environ. Microbiol. **64**:3615-3619.
272. **van Roermund C. W., H. R. Waterham, L. Ijlst, R. J. Wanders.** 2003. Fatty acid metabolism in *Saccharomyces cerevisiae*. Cell. Mol. Life Sci. **60**:1838-1851.
273. **van Roermund C. W., R. Drissen, M. van Den Berg, L. Ijlst, E. H. Hettema, H. F. Tabak, H. R. Waterham, R. J. Wanders.** 2001. Identification of a peroxisomal ATP carrier required for medium-chain fatty acid β-oxidation and normal peroxisome proliferation in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **21**:4321-4329.
274. **Vandesompele J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, F. Speleman.** 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. **3**:RESEARCH0034.
275. **vanKuyk P. A., M. J. de Groot, G. J. Ruijter, R. P. de Vries, J. Visser.** 2001. The *Aspergillus niger* D-xylulose kinase gene is co-expressed with genes encoding arabinan degrading enzymes, and is essential for growth on D-xylose and L-arabinose. Eur. J. Biochem. **268**:5414-5423.
276. **Vashee S., H. Xu, S. A. Johnston, T. Kodadek.** 1993. How do "Zn_2Cys_6" proteins distinguish between similar upstream activation sites? Comparison of the DNA-binding specificity of the GAL4 protein *in vitro* and *in vivo*. J. Biol. Chem. **268**:24699-24706.
277. **Venter J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, and others.** 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304**:66-74.
278. **Walker M. G., W. Volkmuth, E. Sprinzak, D. Hodgson, T. Klingler.** 1999. Prediction of gene function by genome-scale expression analysis: prostate cancer-associated genes. Genome Res. **9**:1198-1203.
279. **Wanders R. J. A., H. R. Waterham.** 2005. Peroxisomal disorders I: biochemistry and genetics of peroxisome biogenesis disorders. Clin. Genet. **67**:107-133.
280. **Wang H., M. Johnston, R. D. Mitra.** 2007. Calling cards for DNA-binding proteins. Genome Res. **17**:1202-1209.
281. **Wang Z., D. M. Soanes, M. J. Kershaw, N. J. Talbot.** 2007. Functional analysis of lipid metabolism in *Magnaporthe grisea* reveals a requirement for peroxisomal fatty acid β-oxidation during appressorium-mediated plant infection. Mol. Plant Microbe Interact. **20**:475-491.
282. **Wanke C., S. Eckert, G. Albrecht, W. van Hartingsveldt, P. J. Punt, C. A. van den Hondel, G. H. Braus.** 1997. The *Aspergillus niger* GCN4 homologue, *cpcA*, is transcriptionally regulated and encodes an unusual leucine zipper. Mol. Microbiol. **23**:23-33.
283. **Watkins P. A., J. F. Lu, S. J. Steinberg, S. J. Gould, K. D. Smith, L. T. Braiterman.** 1998. Disruption of the *Saccharomyces cerevisiae* FAT1 gene decreases very long-chain fatty acyl-CoA synthetase activity and elevates intracellular very long-chain fatty acid concentrations. J. Biol. Chem. **273**:18210-18219.
284. **Wei C., J. Li, R. E. Bumgarner.** 2004. Sample size for detecting differentially expressed genes in microarray experiments. BMC Genomics **5**:87.
285. **White S., M. McIntyre, D. R. Berry, B. McNeil.** 2002. The autolysis of industrial filamentous fungi. Crit. Rev. Biotechnol. **22**:1-14.
286. **Winzeler E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, and others.** 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science **285**:901-906.
287. **Wittenberg C., S. I. Reed.** 2005. Cell cycle-dependent transcription in yeast: promoters, transcription factors, and transcriptomes. Oncogene **24**:2746-2755.
288. **Woese C. R.** 2004. A new biology for a new century. Microbiol. Mol. Biol. Rev. **68**:173-186.

289. **Wolfe C. J., I. S. Kohane, A. J. Butte.** 2005. Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. BMC Bioinformatics **6**:227.
290. **Wolfram S.** 2002. A new kind of science. Wolfram Media, Inc.
291. **Wosten H. A., S. M. Moukha, J. H. Sietsma, J. G. Wessels.** 1991. Localization of growth and secretion of proteins in *Aspergillus niger*. J. Gen. Microbiol. **137**:2017-2023.
292. **Yang Y. H., T. Speed.** 2002. Design issues for cDNA microarray experiments. Nat. Rev. Genet. **3**:579-588.
293. **Young E. T., N. Kacherovsky, K. Van Riper.** 2002. Snf1 protein kinase regulates Adr1 binding to chromatin but not transcription activation. J. Biol. Chem. **277**:38095-38103.
294. **Yuan X., J. A. Roubos, C. A. M. J. J. van den Hondel, A. F. J. Ram.** 2008. Identification of InuR, a new $Zn(II)_2Cys_6$ transcriptional activator involved in the regulation of inulinolytic genes in *Aspergillus niger*. Mol. Genet. Genomics **279**:11-26.
295. **Yuan X., R. M. van der Kaaij, C. A. M. J. J. van den Hondel, P. J. Punt, M. J. E. C. van der Maarel, L. Dijkhuizen, A. F. J. Ram.** 2008. *Aspergillus niger* genome-wide analysis reveals a large number of novel α-glucan acting enzymes with unexpected expression profiles. Mol. Genet. Genomics **279**:545-561.
296. **Zakim D.** 2000. Thermodynamics of fatty acid transfer. J. Membr. Biol. **176**:101-109.
297. **Zar J.** 1996. Biostatistical analysis. Prentice-Hall.

Nederlandse samenvatting

Inleiding

Aspergillus niger is een in de biotechnologie-industrie veelgebruikte schimmel die onder natuurlijke omstandigheden leeft van dood plantenmateriaal. Dit plantenmateriaal bestaat voornamelijk uit suikerpolymeren, maar bevat ook lipiden. Deze lipiden zijn bijvoorbeeld aanwezig als olie in plantenzaden als opslag van energie, als onderdeel van de waslaag die bladeren en fruit beschermt, of als fosfolipiden en galactolipiden die bouwstenen zijn van plantencel-membranen. Het werk dat in dit proefschrift is beschreven is het resultaat van een project dat tot doel had om te onderzoeken hoe de lipiden-afbraak en het lipiden-metabolisme van *A. niger* werkt.

A. niger bezit een verscheidenheid aan extracellulaire enzymen die complexe moleculen buiten de cel kunnen afbreken tot kleinere substraten die kunnen worden opgenomen. In tegenstelling tot enzym-systemen die suikerpolymeren afbreken is er weinig bekend over de afbraak en het metabolisme van lipiden door *A. niger*. De volledige genoom-sequentie van *A. niger* stam CBS 513.88 is bepaald door DSM. Uit analyse van deze genoomsequentie bleek dat alle enzymen die noodzakelijk zijn voor de metabole omzetting van lipiden gecodeerd worden op het genoom. Onder deze enzymen bevinden zich 27 lipases met een secretiesignaal, wat duidt op een functie buiten de cel.

Eerder onderzoek aan andere schimmels suggereerde dat lipases waarschijnlijk op transcriptieniveau gereguleerd worden. Vanwege deze aanwijzingen, en vanwege de beschikbaarheid van *A. niger* DNA microarrays in dit project, is ervoor gekozen om in dit project de transcriptionele respons van deze schimmel op verschillende lipiden te bestuderen.

Een eerste microarray-experiment binnen dit project bracht een tweetal tekortkomingen van de gebruikte experimentele aanpak aan het licht: (i) de reproduceerbaarheid van biologische monsters was laag vanwege de variabele groei in schudkolven, en (ii) de in het experiment gebruikte emulgator had een zodanig groot effect op de transcriptionele respons dat een eventuele olie-specifieke respons hierdoor overschaduwd werd. In dit project is eerst aandacht gegeven aan verbetering van de biologische reproduceerbaarheid; dit is beschreven in **hoofdstuk 2 en 3**. Vervolgens wordt de transcriptionele respons van *A. niger* op de suiker xylose en op verschillende olieën beschreven in **hoofdstuk 3 en 4**. De hier verzamelde microarray gegevens worden in **hoofdstuk 5** gebruikt om globale transcriptie-processen en de hieraan ten grondslag liggende mechanismen te bestuderen. Tenslotte wordt in **hoofdstuk 6** een methode beschreven om de functie van genen die coderen voor niet-gekarakteriseerde eiwitten te bepalen met behulp van microarrays en een specifiek voor dit doel ontworpen transcriptiefactor.

Verbeteren van experimentele opzet

Om de effecten van maatregelen tot het verhogen van de reproduceerbaarheid te meten hebben we gebruik gemaakt van 'quantitative real-time PCR (qPCR)' technologie. Met qPCR is het mogelijk om het transcriptie-niveau van een enkel gen tussen twee of meer condities snel en nauwkeurig te bepalen. Omdat deze methode relatieve verschillen bepaalt, is het noodzakelijk dat qPCR data gecorrigeerd wordt voor experimentele variatie. Hiervoor worden veelal stabiel tot expressie komende genen gebruikt: genen waarvan de transcriptie niet beïnvloed wordt door de experimentele omstandigheden. Deze genen worden endogene ('organisme-eigen') referentie genen genoemd.

In **hoofdstuk 2** staat beschreven hoe, met behulp van data van 48 in het laboratorium voorradige microarrays, 321 kandidaat referentiegenen zijn geselecteerd. Vervolgens is het transcriptie- niveau voor een subset van 11 genen met qPCR bepaald voor de twintig biomassa-monsters die ook gebruikt zijn in het (hierboven beschreven) eerste microarray experiment. Uit analyse van microarray data blijkt dat deze 11 genen een meer stabiel genexpressie-niveau hebben dan een set van traditionele referentie-genen die voor *A. niger* gepubliceerd zijn.

In aanvulling op stabiel tot expressie komende *A. niger* genen is ook een extern referentie-gen getest. Dit externe construct, dat tijdens de cDNA synthese stap toegevoegd moet worden, blijkt stabieler dan de 11 kandidaat referentiegenen te zijn. We concludeerden dat het voor hoge kwaliteit qPCR metingen wenselijk is om het transcriptieniveau te bepalen van zowel van het externe referentie-gen als van enkele van de 11 onderzochte endogene referentie genen.

In **hoofdstuk 3** is de geoptimaliseerde qPCR methode gebruikt om de reproduceerbaarheid van fermentor-gegroeide schimmelcultures te bepalen. Groei in fermentoren is beter te controleren dan groei in schudkolven, onder andere vanwege effectieve beluchting en de mogelijkheid om de zuurgraad constant te houden. In dit groeiexperiment werden 15 identiek gegroeide fermentor-cultures geïnduceerd met 0.1 mM xylose. Met behulp van een hiërarchisch opgezet groei-experiment is de variatie in kaart gebracht die in de verschillende processtappen optreedt. Het transcriptieniveau van acht genen is gemeten met qPCR, en deze data is gebruikt voor een variantiecomponenten-analyse. Uit deze analyse bleek dat 68% van de totale variatie komt door verschillen in dagelijkse handelingen, zoals het bereiden van media, het kweken van sporen, en de opbouw van de fermentoren. De absolute variatie is echter klein.

Biologische interpretaties

Met microarrays is, ook in **hoofdstuk 3**, voor een aantal biomassa-monsters de globale transcriptionele respons op xylose bekeken. De expressie van vierentwintig genen bleek significant te verschillen. Deze genen coderen voor enzymen die betrokken zijn bij de afbraak en verdere omzetting van xylose-bevattende plantpolymeren. Daarnaast coderen enkele genen voor enzymen die betrokken zijn bij de hydrolyse van andere polymeren zoals cellulose. Deze studie bevestigt eerdere waarnemingen dat de schimmel het xylose-signaal interpreteert als een algemeen signaal voor de aanwezigheid van een verscheidenheid aan complexe suikerpolymeren.

De fermentor-gebaseerde experimentele opstelling is gebruikt om de transcriptionele respons van *A. niger* op drie lipiden te bestuderen. Dit werk staat beschreven in **hoofdstuk 4**. Fermentoren werden geïnduceerd met 1 mM van de lipiden olijfolie of de voor galactolipiden verrijkte tarwe-olie. Een derde fermentor werd geïnduceerd met zuivere galactolipiden. Analyse van de microarray data liet zien dat er een specifieke transcriptionele respons op olijfolie en op de tarwe-olie is. Met name genen die coderen voor eiwitten die betrokken zijn bij vetzuur-metabolisme en peroxisoom-aanmaak kwamen hoger tot expressie na inductie. Wel was de transcriptionele respons in de tijd verschillend tussen deze twee oliën. Er werd geen transcriptionele respons op de zuivere galactolipiden waargenomen.

In **hoofdstuk 5** is een analyse beschreven van de globale genexpressieprofielen tussen verschillende groepen condities. Hiervoor zijn gegevens van de microarrays uit hoofdstuk 3 en 4 gebruikt, die als 'milde stress' beschreven worden. Daarnaast is een tweede microarray dataset gebruikt die is gebaseerd op experimenten waarin *A. niger* gegroeid is onder 'hevige stress', zoals de aan- of afwezigheid van een koolstofbron in het kweekmedium. Vervolgens is voor iedere gen-paar combinatie uit een subset van 2,773 genen die evolutionair geconserveerde eiwitten coderen, de mate van correlatie tussen de genexpressie profielen berekend. Dit is uitgevoerd voor zowel de mild verstoorde als de zwaar verstoorde omstandigheden, alsook voor de beide microarray datasets samen.

Op basis van deze correlaties zijn drie 'co-expressie netwerken' gemaakt. De netwerken van de milde stress en het netwerk van de gecombineerde data bevatten 99% identieke gen-paren, terwijl ongeveer 25% van de gen-paren (718 gen-paren) in het mild verstoorde condities-netwerk in het sterk verstoorde condities-netwerk zit.

In deze netwerken zijn modules zichtbaar die bestaan uit genen die onderling een sterk correlerend genexpressieprofiel hebben. Aan een aantal van deze modules kan een hoog-

niveau beschrijving van een biologisch proces gekoppeld worden, zoals betrokken bij 'gen expressie' of 'aminozuur metabolisme'. Voor genen binnen een aantal van deze modules zijn geconserveerde DNA-sequenties in de promoter-regio gevonden. Twee van deze motieven zijn bekende bindingsplaatsen voor transcriptie-factoren die inderdaad betrokken zijn bij de biologische processen van de betreffende modules. Daarnaast is een grote module (ongeveer 600 van 1134 genen, ongeveer 5,500 van 10,311 gen-paren) alleen zichtbaar in het sterk verstoorde condities-netwerk. Deze heterogene module kan niet gekoppeld worden aan een duidelijk biologisch proces.

Hoewel de modules zelf geconserveerd lijken te zijn, zijn de locaties van deze modules binnen het gen co-expressie netwerk en de tussenliggende verbindingen tussen de modules verschillend voor de condities. Het is onduidelijk wat de precieze biologische betekenis van deze observaties is.

Een methode tot studie van transcriptie-factoren

In **hoofdstuk 6** wordt een methode beschreven om te bepalen welke genen door een bepaalde transcriptiefactor gereguleerd zouden kunnen zijn. Transcriptiefactoren zijn DNA-bindende eiwitten die genexpressie specifiek kunnen beïnvloeden. De methode is erop gericht om het DNA-bindingsdomein van de XlnR transcriptiefactor, het zogenaamde Zn_2Cys_6 domein, te vervangen door het Zn_2Cys_6 domein van een andere transcriptiefactor uit dezelfde eiwitfamilie. Hierbij wordt verondersteld dat het geconstrueerde eiwit met het veranderde DNA-bindingsdomein de DNA base-paren zal herkennen die normaal door de tweede, niet bekende, transcriptiefactor herkend worden. Door *A. niger* cellen bloot te stellen aan xylose (de inducer van XlnR) worden nu door deze hybride transcriptiefactor niet meer de XlnR-gereguleerde genen tot expressie gebracht, maar die genen die door de tweede transcriptiefactor gereguleerd worden. Met behulp van microarrays kunnen deze genen vervolgens opgepikt worden, waardoor een koppeling tot stand is gekomen tussen de transcriptiefactor en genen onder diens controle.

Helaas werd slechts een construct dat codeert voor een hybride transcriptiefactor met punt-mutatie verkregen. Tests met dit gemuteerde eiwit lieten niet de gewenste verhoging van genexpressie zien. Vanwege de mutatie was het niet mogelijk om vast te stellen of dit negatieve resultaat door de mutatie veroorzaakt werd of doordat de methode in zijn algemeenheid niet werkt.

Conclusie en vooruitblik

Het werk dat beschreven is in dit proefschrift draagt bij aan de verbetering van experimenteel-biologisch onderzoek aan *Aspergillus niger*. Door het reproduceerbaar kunnen groeien van deze schimmel en door het optimaliseren van technieken om transcriptieprofielen te meten is het mogelijk geworden om ruisvrije, kwalitatief hoogwaardige gegevens over genexpressie te verzamelen. De op deze manier verzamelde gegevens geven inzicht in de respons van deze schimmel op de suiker xylose en op verschillende soorten olie. Daarnaast draagt deze studie bij aan een beter begrip van genexpressie en de daaronderliggende regulatie op het niveau van de cel.

In de toekomst zal meer de nadruk komen te liggen op het gebruik van modellen die de interacties en dynamiek tussen moleculen binnen een organisme (of een deel daarvan, zoals een organel) kunnen beschrijven. Totnutoe kunnen alleen wiskundige modellen deze dynamische interacties goed beschrijven. Zowel voor het maken van deze modellen als voor het valideren ervan zijn hoogwaardige kwantitatieve gegevens nodig. In het niet zo verre verleden waren kwantitatieve gegevens slechts beschikbaar voor een selecte groep model-organismen, zoals *Escherichia coli* en *Saccharomyces cerevisiae*. Door het beschikbaar komen van -omics technieken (zoals DNA microarrays) is het nu mogelijk om ook voor andere dan deze model-organismen gegevens te verzamelen en op basis hiervan modellen te ontwerpen en te toetsen. Dit zal er toe leiden dat meer organismen in groter detail bestudeerd kunnen worden, wat mogelijk leidt tot een beter inzicht in de processen die ten grondslag liggen aan de veelzijdigheid van 'leven'.

Dankwoord
Acknowledgements

Dankwoord

De vijf en een half jaar van mijn promotietraject had ik niet kunnen volhouden zonder de hulp en betrokkenheid van mijn vrienden. Iedereen die geluisterd heeft naar mijn verzuchtingen, die geglimlacht heeft over mijn enthousiasme en mijn ongeduld, die niet begreep maar mij toch mijn gang liet gaan, die in mijn kunnen geloofde als resultaten weer niet positief waren: bedankt! Bedankt ook voor de tijd die jullie mij gegeven hebben; tijd die ik niet in emails, telefoongesprekken, en persoonlijk contact heb gestopt maar in pipetteer-avonden en meet-weekeinden. Het is jullie geduld met mij dat dit proefschrift mede mogelijk gemaakt heeft.

Lieve Helen, jij hebt mij de ruimte gegeven om dit onderzoek te kunnen doen. Dank daarvoor, en voor het vertrouwen dat je altijd in mij hebt gehad. Ik ben blij met je eerlijke mening, je gave om te luisteren, en voor het in perspectief plaatsen van mijn enthousiasme en frustraties. Lieve ouders, dank voor een zorgeloze jeugd die als basis voor deze promotie gediend heeft. Schoonouders, broers en zussen: jullie hebben me verwarmd met betrokkenheid en aandacht.

Ik ben iedereen van Fungal Genomics erkentelijk voor hun samenwerking en collegialiteit, ook tijdens de soms roerige tijden binnen de groep. Luisa en Kelly, dank voor de gezellige sfeer en jullie hartelijke lach op kamer 1044. Elena, dank voor prettige gesprekken over niets en alles, ook 's avonds. Miguel, taaldanser, goed wetenschapper, en fijne collega: dank voor harmonie. Bas, ik glimlach nog steeds om je rake opmerkingen. Willy, je ervaring en oog voor detail hebben aan dit proefschrift een duidelijke bijdrage geleverd. Peter, je bio-informatica gereedschap was een waardevolle aanvulling op de biologie in dit project. Harry, Jac, Hans, Geja, René, Henk-Jan, Harrie, Jorg, Christina, Rob, Cozmina, Saïd, Ab en Johan: ieder van jullie heeft op zijn eigen wijze bijgedragen aan mijn ontwikkeling en aan dit proefschrift. Dank jullie wel.

Als begeleider heb ik van studenten zeker meer geleerd dan zij van mij. Nardy, jouw werk heeft mij op de weg van goed experimenteel ontwerp geleid. Canan, je doorzettingsvermogen heeft in slechts enkele maanden gekloneerde cutinases en een gedeeltelijk functionele *A. niger* colony-PCR opgeleverd. Alper, altijd enthousiast; wellicht herken je wat van je werk in hoofdstuk 6. Betül, je bescheidenheid verbergt een krachtig persoon en een goed wetenschapper. Dank voor het knock-out systeem, en je vertrouwen. Mariaan, dank voor je inzet en werk aan standaardisering en aan de

A translation in English is on page 172.

overexpressie-vectoren. Violeta, dank voor je enthousiasme en voor het werk aan de chimere transcriptiefactor.

Binnen de diversiteit van Microbiologie (eerst in 'het andere gebouw', later 'samen') heb ik veel aardige en hulpvaardige collega's leren kennen. Dank voor jullie gastvrijheid en jullie bereidheid mee te denken over praktische problemen. Wim, met name jij bedankt voor hulp met computers en fermentoren. De collega's van Humane Voeding op de begane grond van gebouw 'De Valk' wil ik bedanken voor hun toegankelijkkeid en openheid. Jullie hulp bij qPCR en microarray technologiën was welkom. Heleen, ons maandag-diner was een ijkpunt in mijn week, dank voor leuke gesprekken over wetenschap en meer. Robert, dank voor je sterke vriendschap en je luisterend oor. Ik ben blij met het mooie resultaat van onze samenwerking. Gabriele, uit de Japan-reis vloeide een vriendschap voort waar ik erg blij mee ben. Machtelt, Mariët en Peter, ik dank jullie voor de prettige samenwerking voor hoofdstuk 5.

DSM Food Specialties heeft een grote rol gespeeld bij de totstandkoming van dit proefschrift, onder andere door het beschikbaarstellen van operationele kennis en toegang tot DNA microarrays. Rutger en Rob, dit project is mede dankzij jullie steun voorspoedig van start gegaan. Jullie en alle leden van dit Algebra-project dank ik voor hun bijdragen en commentaar.

John, ik ben blij dat je op mijn verzoek om mijn promotor te worden bent ingegaan. Instappen aan het eind van mijn promotietraject was lastig omdat het experimentele werk al gedaan was. Jouw inzet en constructieve bijdragen hebben dit proefschrift versterkt. Ik wil je bedanken voor een prettige samenwerking en motiverende gesprekjes.

Leo, dit proefschrift draagt jouw stempel. Door jouw afscherming kon ik mij richten op de wetenschappelijke inhoud van dit project, en je optimisme heeft mij door menig lastige periode gehaald. Ik waardeer je duidelijkheid en je bereidheid om niet de makkelijkste maar de best mógelijke weg te volgen. Je stelt je open voor andere mensen en meningen en bent bereid om je oordeel bij te stellen – wellicht de essentie van een goede wetenschapper. Daarnaast wil ik je bedanken voor het prettige persoonlijk contact, de gesprekken in Loburg, het goede eten. Dank voor dat alles, en voor al je vertrouwen in mij.

Douwe

Acknowledgements

Without the involvement and support of my friends during these five and a half years I could not have kept taking on this doctoral degree. To all who has listened to my lamenting, who smiled at my enthusiasm and impatience, who did not understand but nevertheless allowed me to continue, who had faith in my skills when once again results were not positive: thank you. In addition, thank you for your gift of time; time that was not spent on e-mails, telephone conversations, and personal contact, but rather on evening pipetting sessions and laboratory weekends. Your patience with me has made this Ph.D. thesis possible too.

Dear Helen, you have given me the freedom to pursue this research. Thank you for that, and for your confidence and trust. I am pleased by your honest opinions, your gift of listening, and by your ability to place into perspective my enthusiasm and frustrations. Dear parents, thank you for a carefree youth that has served as solid base persuing this degree. Parents-in-law, brothers and sisters: your concern and attentiveness have warmed me.

I am grateful to past and current members of the Fungal Genomics group for their co-operation and fellowship, even in sometimes tempestuous periods within the group. Luisa and Kelly, thank you for an enjoyable and cosy atmosphere in room 1044, and for your hearty laughs. Elena, thank you for pleasant conversations about all things and nothing, also during evenings. Miguel, languagedancer, good scientist, and fine colleague: thank you for balance. Bas, your apt remarks still make me smile. Willy, your experience and sharp eye for detail contributed heavily to this thesis. Peter, your bio-informatics tools were a valuable addition to this project's biology. Harry, Jac, Hans, Geja, René, Henk-Jan, Harrie, Jorg, Christina, Rob, Cozmina, Saïd, Ab and Johan: each of you has contributed to my development and to this thesis in his or her own way. Thank you for that.

Being a supervisor, I am sure I have learned more from my students than they learned from me. Nardy, your work lead me to the path of experimental design. Canan, in just a few months your persistence resulted in cloned cutinases and a partially functioning *A. niger* colony-PCR method. Alper, always enthousiastic: you might recognize some of your work in the sixth chapter. Betül, you hide a strong personality and an excellent scientist behind modesty. Thank you for the knock-out system, and your confidence. Mariaan, thank you for your dedication and efforts to standardization and to the over-expression vectors. Violeta, thans for enthousiasm and your work on the chimeric transcription factor.

Within the diverse Microbiology group (at first in 'the other building', later 'together') I came to meet many nice and helpful colleagues. Thank you for hospitality and your willingness to reflect upon my practical problems. Wim, you especially I would like to thank for help with computers and fermentors. I would like to thank the Human Nutrition colleagues at the 'De Valk' building's ground floor for accessibility and openness. Your help at qPCR and microarray technologies was welcomed. Heleen, thanks for conversations about science and non-science during our weekly Monday-dinner. Robert, thanks a lot for strong friendschip and for listening. I really appreciate the result of our mutual co-operation. Gabriele, I am grateful for our friendship that arose out of the Japan-trip. Machtelt, Mariët, and Peter, I would like to thank you for your pleasant co-operation for the work on the fifth chapter.

DSM Food Specialties had great impact on the finalization of this thesis, among other things by providing operational knowledge and access to their DNA microarray platform. Rutger and Rob, the project went off well thanks also to your efforts. I thank you and all members of the Algebra project for their contributions.

John, I am glad that you accepted my request to become thesis supervisor. Given that all experimental work had been finished, to join at the end of this trajectory was not an easy thing. Your efforts and constructive contributions have strenghtened this thesis. I would like to thank you for a fine co-operation and motivating chats.

Leo, this thesis bears your signature. I was able to concentrate on the scientific content of this project due to your guarding, and your optimistic views brought me through many an arduous period. I appreciate your clarity and your willingness to not take the easy road but instead choose the best possible road. You open your mind to other views and other persons and you have the willingness to adjust your judgements – perhaps, the essence of being a good scientist. Beside this, I would like to thank you for the pleasant personal contact, the chats at Loburg, and the excellent food. Thanks for all of that, and for your confidence in me.

About the author

Douwe van der Veen was born on the 27th of August, 1976, in the city of Utrecht, The Netherlands. He completed his secondary eduction at Sprengeloo Christelijke Scholengemeenschap in Apeldoorn in 1995. In the same year, he began his tertiary education in Bioprocesstechnology at Wageningen University. Within this multidisciplinary M.Sc. course, Douwe chose the 'cellular-molecular' specialization.

Douwe's interest is in regulatory mechanisms and signal transduction cascades, with the emphasis on these processes in microorganisms. His academic internship was at the Molecular Genetics of Industrial Microorganisms group (supervisors Dr. A.A. Hasper and Dr. ir. L.H. de Graaff). Here, Douwe worked on the analysis of functional domains of the transcriptional activator XlnR of *Aspergillus niger*.

In the first of two research internships, Douwe investigated the nature of protein-protein interactions of KinA, the major kinase of the sporulation phosphorelay of *Bacillus subtilis*. This research was conducted at the Hoch Lab of The Scripps Research Institute, La Jolla CA, USA (supervisors Dr. K. Stephenson, Dr. M. Perego, Dr. J.A. Hoch). The second internship was at Roche Vitamins, Basel (Switzerland) where Douwe investigated regulatory mechanisms of the pantothenate biosynthesis pathway of *B. subtilis* (supervisors Dr. G.R. Schyns, Dr. J.B. Perkins).

After successful completion of his M.Sc. Bioprocesstechnology in 2003, Douwe started on a Ph.D. project entitled "ALGeBRa: *Aspergillus* lipase genomics for bakery research", the results and outcomes of which are described in this thesis. This thesis' work was conducted within the Fungal Genomics group, Wageningen University, and was supervised by Dr. L.H. de Graaff.

VLAG graduate school activities

Discipline specific activities

RT-PCR workshop series, Free University Amsterdam, 2004
Affymetrix workshop, Affymetrix, Wageningen, 2004
Systems Biology: principles of *~omics* data analysis, NBIC, Nijmegen, 2005
Genedata workshop, Delft University, 2005

ALW Platform Molecular Genetics, Lunteren, 2005 ¶
11th Netherlands Biotechnology Congress, Ede, 2006 ‡
3d Kluyver Centre Symposium, Noordwijkerhout, 2006 ‡
4th *Aspergillus* meeting, Asilomar CA, USA, 2007 ‡
24th Fungal Genetics Conference, Asilomar CA, USA, 2007 ‡
13th European Congress on Biotechnology, Barcelona, Spain, 2007 ‡ ¶
12th Netherlands Biotechnology Congress, Ede, 2008 ‡

General courses

VLAG PhD week, 2004
Introductory course on didactics, Wageningen University, 2004-05
Scientific writing, Wageningen University, 2007

Optional courses and activities

Preparation PhD research proposal, 2003
PhD study trip Japan, Lab. of Microbiology, 2004
Lab. of Microbiology biweekly group meetings, Wageningen, 2003-08
Section Fungal Genomics biweekly meetings, Wageningen, 2003-08

---

‡ Poster
¶ Presentation

Colophon

This thesis is printed by GVO drukkers & vormgevers B.V. | Ponsen & Looijen

Text: 9.5 / 13.0 pt Gentium. The Gentium font family is under the SIL Open Font License, http://scripts.sil.org/Gentium.
Headers: 10.0; 14.0; 24 pt `DejaVu Sans Mono`. The DejaVu font family is © Bitstream Inc. and released in the public domain.

Inkscape 0.46, http://www.inkscape.org
GIMP 2.6, http://www.gimp.org
Blender 2.48a/svn, http://www.blender.org
Scribus 1.3.5svn, http://www.scribus.net